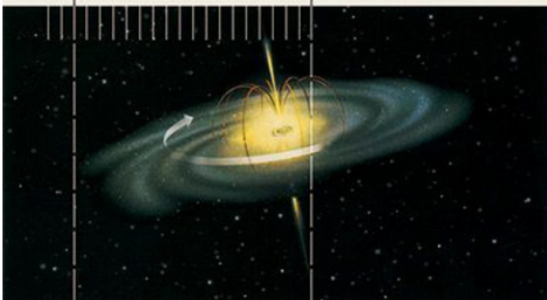
宇宙系列



# A Brief History of Time

# 时间简史

[英] 史蒂芬·霍金 / 著　　许明贤　吴忠超 / 译

第一推动

湖南科学技术出版社

# 版权信息

第一推动·探索宇宙科学指南

作者：【英】霍金　【美】基普·S·索恩　【美】布赖恩·格林　【英】柯文尼　【英】戴维斯

湖南科学技术出版社
HUNAN SCIENCE & TECHNOLOGY PRESS

# 目录


版权信息
时间简史
总序
译者序
《时间简史》首版导言
前言
第一章 我们的宇宙图象
我们的宇宙图象
第二章 空间和时间
空间和时间
第三章 膨胀的宇宙
膨胀的宇宙
第四章 不确定性原理
不确定性原理
第五章 基本粒子和自然的力
基本粒子和自然的力
第六章 黑洞
黑洞
第七章 黑洞不是这么黑的
黑洞不是这么黑的
第八章 宇宙的起源和命运
宇宙的起源和命运
第九章 时间箭头
时间箭头
第十章 虫洞和时间旅行
虫洞和时间旅行

第十一章 物理学的统一
物理学的统一
第十二章 结论
结论
阿尔伯特·爱因斯坦
伽利略·伽利雷
艾萨克·牛顿
小辞典
感谢
第一推动丛书书目
黑洞与时间弯曲
总序
序
前言
导引
序幕：黑洞之旅
第1章 空间和时间的相对论
第2章 空间和时间的卷曲
第3章 黑洞，发现与拒绝
第4章 白矮星之谜
第5章 坍缩是必然的
第6章 坍缩成什么？
第7章 黄金年代
第8章 寻找
第9章 意外发现
第10章 曲率波
第11章 实在是什么？
第12章 黑洞蒸发
第13章 洞里
第14章 虫洞和时间机器

尾声
致谢
人物
年表
名词
注释
文献目录
译后记
在黑洞的地平线上——重印后记
重印后记
宇宙的琴弦
总序
序
为B·格林《宇宙的琴弦》喝彩
第一篇 知识的边缘
第二篇 空间、时间和量子的困境
第三篇 宇宙交响曲
第四篇 弦理论与时空结构
第五篇 21世纪的统一
科学名词解释
推荐读物
译后记
重印后记
宇宙的结构
总序
第1部分 实在性之舞台
第2部分 时间与经验
第3部分 时空与宇宙学
第4部分 起源与统一
第5部分 真实与想象

注释
术语表（以拼音为序）
推荐书目
时间之箭
总序
前言
序
第一章 时间的形象
第二章 牛顿物理学的兴起：时间失去了方向
第三章 时间使爱因斯坦受挫
第四章 时间的量子跃迁
第五章 时间之箭：热力学
第六章 创造性演化
第七章 时间之箭与生命之箭
第八章 统一的时间景象
第九章 未结束的探索
上帝与新物理学
总序
致谢
前言
第一章 科学与宗教在变化着的世界中
第二章 创世
第三章 上帝创造了宇宙？
第四章 为什么会有个宇宙？
第五章 什么是生命？整体论对还原论
第六章 精神与灵魂
第七章 自我
第八章 量子因素
第九章 时间
第十章 自由意志与决定论

第十一章 物质的基本结构
第十二章 偶然还是设计？
第十三章 黑洞与宇宙混沌
第十四章 奇迹
第十五章 宇宙的终结
第十六章 宇宙是“免费午餐”吗？
第十七章 物理学家心目中的大自然
精选参考书目
注释
名词对照表
译后记
物理学的困惑
总序
绪言
致谢
第一篇 未完成的革命
第二篇 弦论简史
第三篇 弦论之外
第四篇 向经验学习
译后记

# 时间简史

# A BRIEF HISTORY OF TIME

UPDATED AND EXPANDED
TENTH ANNIVERSARY EDITION

STEPHEN HAWKING

BANTAM BOOKS

NEW YORK LONDON TORONTO SYDNEY AUCKLAND

**A BRIEF HISTORY OF TIME**

**A Bantam Book**

**Publishing History**

**Bantam illustrated hardcover edition published November 1996**

**Bantam hardcover edition/September 1998**

**Bantam trade paperback edition/September 1998**

**All rights reserved.**

**Copyright © 1988, 1996 by Stephen Hawking**

**Illustrations copyright © 1988 by Ron Miller**

**Front cover photo of author copyright © 1988 by David Montgomery**

**Front cover universe background copyright © 1988 by Roger Ressmeyer**

**Book design by Glen M. Edelstein**

**No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without permission in writing from the publisher.**

**For information address: Bantam Books.**

Library of Congress Cataloging-in-Publication Data Hawking, S. W. (Stephen W.)

A brief history of time / Stephen Hawking.—Updated and expanded tenth anniversary ed.

p. cm.

Includes index.

ISBN 0-553-10953-7.—ISBN 0-553-38016-8

1. Cosmology. I. Title.

QB981.H377 1998

523.1—dc21 98-21874

CIP

Published simultaneously in the United States and Canada

Bantam Books are published by Bantam Books, a division of Bantam Doubleday Dell

Publishing Group, Inc. Its trademark consisting of the words "Bantam Books" and the portrayal of a rooster, is Registered in U.S. Patent and Trademark Office and in other countries. Marca Registrada. Bantam Books, 1540 Broadway, New York, New York 10036.

PRINTED IN THE UNITED STATES OF AMERICA

FFG 20 19 18 17 16 15 14 13

# 总序

科学，特别是自然科学，最重要的目标之一，就是追寻科学本身的原动力，或曰追寻其第一推动。同时，科学的这种追求精神本身，又成为社会发展和人类进步的一种最基本的推动。

科学总是寻求发现和了解客观世界的新现象，研究和掌握新规律，总是在不懈地追求真理。科学是认真的、严谨的、实事求是的，同时，科学又是创造的。科学的最基本态度之一就是疑问，科学的最基本精神之一就是批判。

的确，科学活动，特别是自然科学活动，比较起其他的人类活动来，其最基本特征就是不断进步。哪怕在其他方面倒退的时候，科学却总是进步着，即使是缓慢而艰难地进步，这表明，自然科学活动中包含着人类的最进步因素。

正是在这个意义上，科学堪称为人类进步的“第一推动”。

科学教育，特别是自然科学的教育，是提高人们素质的重要因素，是现代教育的一个核心。科学教育不仅使人获得生活和工作所需的知识和技能，更重要的是使人获得科学思想、科学精神、科学态度以及科学方法的熏陶和培养，使人获得非生物本能的智慧，获得非与生俱来的灵魂。可以这样说，没有科学的“教育”，只是培养信仰，而不是教育。没有受过科学教育的人，只能称为受过训练，而非受过教育。

正是在这个意义上，科学堪称为使人进化为现代人的“第一推

动”。

近百年来，无数仁人智士意识到，强国富民再造中国离不开科学技术，他们为摆脱愚昧与无知做了艰苦卓绝的奋斗，中国的科学先贤们代代相传，不遗余力地为中国的进步献身于科学启蒙运动，以图完成国人的强国梦。然而应该说，这个目标远未达到。今日的中国需要新的科学启蒙，需要现代科学教育。只有全社会的人具备较高的科学素质，以科学的精神和思想、科学的态度和方法作为探讨和解决各类问题的共同基础和出发点，社会才能更好地向前发展和进步。因此，中国的进步离不开科学，是毋庸置疑的。

正是在这个意义上，似乎可以说，科学已被公认是中国进步所必不可少的推动。

然而，这并不意味着，科学的精神也同样地被公认和接受。虽然，科学已渗透到社会的各个领域和层面，科学的价值和地位也更高了，但是，毋庸讳言，在一定的范围内，或某些特定时候，人们只是承认“科学是有用的”，只停留在对科学所带来的后果的接受和承认，而不是对科学的原动力，科学的精神的接受和承认。此种现象的存在也是不能忽视的。

科学的精神之一，是它自身就是自身的“第一推动”。也就是说，科学活动在原则上是不隶属于服务于神学的，不隶属于服务于儒学的，科学活动在原则上也不隶属于服务于任何哲学。科学是超越宗教差别的，超越民族差别的，超越党派差别的，超越文化的地域差别的，科学是普适的、独立的，它自身就是自身的主宰。

湖南科学技术出版社精选了一批关于科学思想和科学精神的世界名著，请有关学者译成中文出版，其目的就是为了传播科学的精神，科学的思想，特别是自然科学的精神和思想，从而起到倡导科学精神，推动科技发展，对全民进行新的科学启蒙和科学教育的作用，为中国的进步做一点推动。丛书定名为《第一推动》，当然并非说其中

每一册都是第一推动，但是可以肯定，蕴含在每一册中的科学的内容、观点、思想和精神，都会使你或多或少地更接近第一推动，或多或少地发现，自身如何成为自身的主宰。

《第一推动》丛书编委会

# 译者序

宇宙学是一门既古老又年轻的学科。作为宇宙中高等生物的人类不满足于自身的生存和种族的绵延，还一代代地探索着存在和生命的意义。但是，人类理念的进化是极其缓慢和艰苦的。从亚里士多德-托勒密地心说到哥白尼-伽利略日心说的演化就花了大约2000年的时间。令人吃惊的是，尽管人们知道世间的一切都在运动，只有到了20世纪20年代哈勃发现了红移定律后，宇宙演化的观念才进入人类的意识。在此之前，人们甚至从未想到过宇宙还会演化。无论是牛顿的万有引力理论还是爱因斯坦的广义相对论都不能得到稳态的宇宙模型。为了得到一个这样的模型，爱因斯坦甚至不惜牺牲理论的美丽，将宇宙常数引进他的方程。可见宇宙演化的观念并非产生于这些天才的头脑之中。

哈勃的发现标志着现代宇宙学的诞生。他的红移定律说，从星系光谱的红移可以推断，越远的星系以越快的速度飞离开我们，这表明整个宇宙处于膨胀的状态。从时间上倒溯到过去，估计在100亿到200亿年之前发生过一次开天辟地的大爆炸，宇宙就从这个极其致密极热的状态中诞生。伽莫夫在1948年发表的关于热大爆炸模型的文章中作出了一个惊人的预言，早期大爆炸的辐射仍残存在我们的周围，不过由于宇宙膨胀引起的红移，其绝对温度只余下几度了。在这种温度下，辐射处于微波的波段。然而，在1965年彭齐亚斯和威尔逊观测到宇宙微波背景辐射之前，人们并不认真对待这个预言。

一般认为，爱因斯坦的广义相对论是描述宇宙的正确理论。在经典广义相对论的框架中，霍金和彭罗斯，在很一般的条件下，证明了

时空一定存在奇点，最著名的奇点即是黑洞里的奇点和宇宙大爆炸处的奇点。所有定律和可预见性都在奇点处失效。奇点可以看作时空的边缘或边界。只有给定了奇点处的边界条件，才能从爱因斯坦方程得出宇宙的演化。由于边界条件只能由宇宙外的造物主给定，所以宇宙的命运就操纵在造物主手中。这就是从牛顿时代起一直困扰人类智慧的第一推动问题。

如果时空没有边界，则就不必劳驾上帝进行第一推动了。这只有在量子引力论中才能做到。霍金认为宇宙的量子态是处于一个基态，而时空可被看成是一个有限无界的四维面，正如地球的表面一样，只不过多了两个维数而已。宇宙中的所有结构都起源于量子力学的不确定性原理允许的最小起伏。从一些简单的模型计算可得出和天文观测相一致的推论，如星系团、星系、恒星等成团结构，宇宙大尺度的均匀性和各向同性，时空的平性，时空的维数，太初引力波和太初黑洞，以及时间的箭头等。霍金的量子宇宙学在于它真正使宇宙论成为一门成熟的科学。它是一个自足的理论，即在原则上，单凭科学定律我们便可以将宇宙中的一切都预言出来。

本书作者是当代最重要的广义相对论家和宇宙学家。

20世纪70年代他和彭罗斯一道证明了著名的奇点定理。之后他还证明了黑洞的面积定理，即随着时间增加黑洞的表面积不减。这很自然地使人将这面积和热力学的熵联想起来。

1973年，他考虑黑洞附近的量子效应，发现黑洞会像黑体那样发出辐射。其辐射的温度和黑洞质量成反比，这样黑洞就会因为辐射而慢慢变小，而温度却越变越高，它以最后一刻的爆炸而告终。黑洞辐射的发现具有极其基本的意义，它将引力、量子力学和热力学统一在一起。

1974年以后，他的研究转向量子引力论。他利用费恩曼的对历史求和方法，自然地处理时空的非平凡的拓扑效应，开创了引力热力

学。

1980年，他的兴趣转向量子宇宙学，研究宇宙的无中生有的创生机制，企图一劳永逸地解决第一推动问题。

霍金的生平是非常富有传奇性的。在科学成就上，他是有史以来最杰出的科学家之一，而他的贡献是在他20年之久被肌萎缩性（脊椎）侧索硬化症禁锢在轮椅上的情形下做出的，这真正是空前的。因为他的贡献对于人类的观念有深远的影响，所以媒体对他早已广为报道。尽管如此，译者之一于1979年第一回见到他时的情景至今还历历在目。那是第一次参加他领导的小组的讨论班时，门打开后，忽然脑后响起一种非常微弱的电器的声音，回头一看，只见一个骨瘦如柴的人斜躺在电动轮椅上，他自己驱动着电开关。译者尽量保持礼貌而不显出过分吃惊，但是他对首次见到他的人对其残废程度的吃惊早已习惯。他要很费劲才能抬起头来。在失声之前，他只能用非常微弱的变形的语言交流，这种语言只有在陪他工作、生活几个月后才能通晓。他不能写字，看书必须依赖于一种翻书页的机器，读文献时必须让人将每一页摊平在一张大办公桌上，然后他驱动轮椅如蚕吃桑叶般地逐页阅读。人们不得不对人类中居然有以这般坚强意志追求终极真理的灵魂从内心产生深深的敬意。他每天必须驱动轮椅从他的家——剑桥西路5号，经过美丽的剑河、古老的国王学院驶到银街的应用数学和理论物理系的办公室。该系为了他的轮椅行走方便特地修了一段斜坡。

在富有学术传统的剑桥大学，他目前担任着也许是有史以来最为崇高的教授职务，那是牛顿和狄拉克担任过的卢卡斯数学教授。

本书译者之一曾受教于霍金达4年之久，并在它的指导下完成了博士论文。从他对译者私事的帮助可以体会到，他是一位富有人情味的人。此书即是受霍金之托而译成中文，以供人类1/5的人口了解他的学说。

许明贤　吴忠超

# 《时间简史》首版导言

我们在几乎对世界毫无了解的情形下进行日常生活。我们对于使生命得以实现的阳光的产生机制，对于将我们束缚在地球上，否则我们就会以涡旋的轨道被抛到太空去的重力，对于我们由之构成并依赖其稳定性的原子思考得很少。除了小孩（他们知道太少，会不知轻重地问重要的问题），我们中很少人会用大量时间惊讶自然界为何这个样子；宇宙从何而来或它是否总在这儿；时间会不会有朝一日倒流，并因此导致果先于因；或者人类认识是否有一最终的极限。甚至我曾遇到一些小孩，他们想要知道黑洞是什么样的？物质的最小的部份是什么？为何我们记住过去而不是将来；如果早先是紊乱的，则今天显然是有序的，这究竟是怎么回事？为何存在一个宇宙？

在我们社会里，父母或老师仍然依惯例用耸肩膀或借助模糊回想起的宗教格言去回答这些问题的大部份。有一些人则对这一类的问题感到不舒服，因为它们如此生动地暴露了人类理解的局限性。

但是，哲学和科学的大部份即是由这种好奇心所驱动的。越来越多的成年人愿意问这类问题，并且他们偶尔得到一些使其惊奇的答案。我们这些离开原子和恒星同样远的人类，正在扩大自己探索的视野去拥抱这非常小和非常大的对象。

1974年初，大约在海盗空间飞船登陆火星之前两年，我参加在英国由伦敦皇家学会主办的关于探索如何寻找天外生命的会议。

在会议中间休息时，我注意到在隔壁的大厅里正举行一个更大得多的会议，出于好奇心我进去了。我很快意识到自己见证了一个古代的仪式，是一个新会员参加皇家学会——这个本行星上最古老的学术组织的授职式。前排一位在轮椅中的年轻人正非常缓慢地将他的名字签在一本书上，而这本书的最前页是伊萨克·牛顿的签名。当他最后签好时，大厅里响起了一阵响亮的掌声。史蒂芬·霍金，甚至在那时就是一位传说中的人物。

现在霍金是剑桥大学的卢卡逊数学教授。这个职务曾为牛顿，后来又为狄拉克，这两位非常大和非常小的世界的有名的探索者担任过。他是他们的毫不逊色的继承人。这本霍金首次为非专家写的书，会给外行读者以多种类的酬劳。和这本书的广泛的内容一样有趣的是对作者智力工作的浏览。物理、天文、宇宙学和勇气的前沿被清晰地呈现在本书之中。

这又是一本关于上帝……或许是关于上帝不存在的书。处处充满了上帝这个字眼。霍金着手回答爱因斯坦著名的关于上帝在创生宇宙时有无选择性的问题。正如霍金明白声称的，他企图要去理解上帝的精神。这使得迄今所有努力的结论更加出人意外：一个空间上无边缘、时间上无始无终、并且造物主无所事事的宇宙。

——卡尔·萨根

康奈尔大学　纽约州

# 前言

我没有为《时间简史》的原版写前言。那是卡尔·萨根写的。我写了简短的“感谢”，有人建议我感谢每一个人。但是有些支持过我的基金会不甚高兴，由于被我提及而收到大量的申请。

我认为没有任何人，我的出版者，代理人，甚至我自己曾预料到这本书会这么畅销。它荣登伦敦《星期日时报》畅销书榜237周，这比任何书都久（显然，《圣经》和莎士比亚不算在内）。它被译成40来种文字，在全世界每750人都拥有一册，包括男人、妇女和儿童。正如纳珍·米尔伏德（我的前博士后）评论的：我的物理著作比麦当娜谈性的书还更好卖。

《时间简史》的成功表明，人们对于重大问题有广泛的兴趣：诸如我们从何而来？宇宙为何是这样子的？

我已趁此机会更新本书，并将从首版（1988年4月愚人节）以来理论和观测的新结果纳入。我新添了虫洞和时间旅行一章。爱因斯坦的广义相对论为我们提供了创生和维持虫洞的可能性，那是连接时空中不同区域的细管。如是，我们也许可以利用它们来进行星系之间快速旅行或在时间中旅行到过去。当然，我们从未邂逅到来自未来的人（也许我们曾经有过？）。对此我将给出一种可能的解释。

我还描述了今年在寻求“对偶性”或显然不同的物理理论之间的对应方面的进展。这些对应强烈地表明，存在一种完备的统一物理理论，但是它们也暗示，也许不可能用一个单独表述来表达这个理论。相反，在不同的情形下，我们必须使用基本理论的不同影像。这和描

绘地球表面很相似，人们不能只用一张单独的地图，在不同的区域必须用不同的地图。这就变革了我们的科学定律的统一观，但是他并没有改变最重要的一点：一组我们能够发现并理解的合理的定律制约着宇宙。

在观测方面，迄今最主要的发展是由COBE（宇宙背景探险者）和其他合作者测量的宇宙微波背景辐射中的起伏。这些起伏是创生的指纹，这些在光滑均匀的早期宇宙上的微小的初始无规性后来成长为星系、恒星以及在我们周围看到的所有结构。起伏的形式和无边界设想的预言相吻合。无边界设想说，宇宙在虚时间方向没有边界或边缘。为了区分这个设想和对背景中的起伏的其他可能的解释，还需要进一步的观测。然而，在几年之内，我们就应能知道，我们能否相信自己生活在一个完全自足的无始无终的宇宙之中。

史蒂芬·霍金

1996年5月剑桥

# 第一章 我们的宇宙图象

一位著名的科学家（据说是伯特兰·罗素）曾经作过一次天文学讲演。他描述了地球如何围绕着太阳公转，而太阳又是如何围绕着称之为我们星系的巨大的恒星集团的中心公转。演讲结束之际，坐在房间后排的一位小个老妇人起立说道：“你讲的是一派胡言。实际上，世界是驮在一只巨大乌龟背上的平板。”这位科学家露出高傲的微笑，然后答道：“那么这只乌龟站在什么上面的呢？”“你很聪明，年轻人，的确很聪明”，老妇人说，“不过，这是一只驮着一只，一直驮下去的乌龟塔啊！”

大多数人会觉得，把我们的宇宙喻为一个无限的乌龟塔相当荒谬。但是我们凭什么就自认为知道得更好呢？我们对宇宙了解了多少？而我们又是如何知道的呢？宇宙从何而来，又将向何处去？宇宙有开端吗？如果有的话，在开端之前发生了什么？时间的本质是什么？它会有一个终结吗？物理学中最近的突破，使我们有可能为其中一些长期以来悬而未决的问题提供答案，而奇妙的新技术是实现这些突破的部分原因。对我们而言，这些答案也许有朝一日会变得和地球围绕着太阳公转那么显而易见——或许也会变得和乌龟塔一样荒谬，只有时间（不管其含义如何）才能裁决。

早在公元前340年，希腊哲学家亚里士多德在他的《论天》一书中，就能够对于地球是一个圆球而不是一块平板这个信念提出两个有力的论证。第一，他意识到，月食是由于地球运行到太阳与月亮之间引起的。地球在月亮上的影子总是圆的，这只有在地球本身为球形的

前提下才成立。如果地球是一块平坦的圆盘，除非月食总是发生在太阳正好位于这个圆盘中心的正下方的时刻，否则地球的影子就会被拉长而成为椭圆形。第二，希腊人从旅行中知道，在南方地区观测北极星，比在较北地区，北极星在天空中显得较低。（由于北极星位于北极的正上方，所以它出现在北极的观察者的头顶上，而对于赤道上的某观察者，北极星刚好出现在地平线上。）

根据北极星在埃及和在希腊表观位置的差别，亚里士多德甚至估计出地球大圆长度为400000斯特迪亚。现在不能准确地知道，1斯特迪亚的长度究竟是多少，但也许是200码（1码＝0.9144米）左右，这样就使得亚里士多德的估计大约为现在接受数值的两倍。希腊人甚至为地球是球形提供了第三个论证，否则何以从地平线驶来的船总是先露出船帆，然后才露出船身？

亚里士多德认为地球是不动的，太阳、月亮、行星和恒星都以圆周为轨道围绕着地球公转。他相信这些，是因为他认为地球是宇宙的中心，而圆周运动是最完美的；他的这种看法是基于某些神秘的原因。公元2世纪，这个思想被托勒密精制成一个完整的宇宙学模型。地球处于正中心，8个天球包围着它，这8个天球分别负载着月亮、太阳、恒星和5个当时已知的行星：水星、金星、火星、木星和土星（图1.1）。为了说明在天空中观察到的这些行星的相当复杂的轨道，人们认为它们本身沿着附在相应天球上的更小的圆周运动。最外层的天球携带着所谓的固定恒星，它们的相对位置总是保持不变，但是总体围绕着天空旋转。最后一层天球之外为何物一直不很清楚，但是它肯定不是人类所能观测到的宇宙部分。

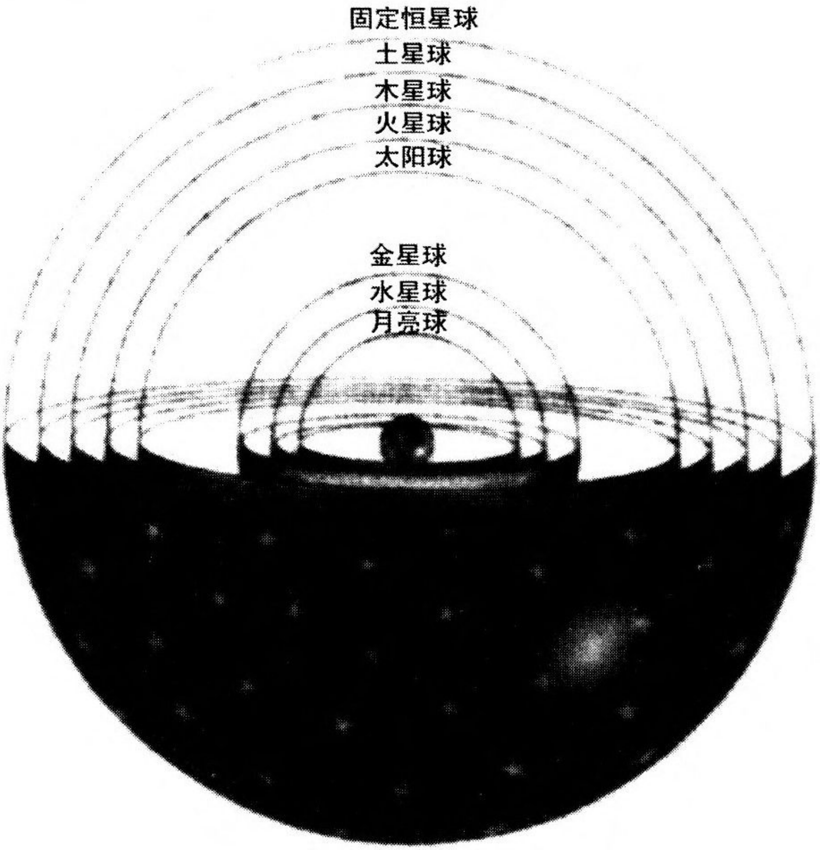



图1.1　从最里面往外面顺序为月亮球、水星球、金星球、太阳球、火星球、木星球、土星球和固定恒星球。最中心为地球。

托勒密模型的系统可以相当精密地预言天体在天空中的位置。但是为了正确地预言这些位置，托勒密不得不假定，月亮遵循的轨道有时使它离地球的距离是其他时候的一半。这意味着月亮有时显得要比其他时候大一倍！托勒密承认这个瑕疵，但是尽管如此，他的模型被广泛地，虽然不是普适地接受。它被基督教会接纳为与《圣经》相一致的宇宙图象。这是因为它具有巨大优势，即在固定恒星天球之外为天堂和地狱留下了大量的空间。

然而，1514年波兰教士尼古拉·哥白尼提出了一个更简单的模

型。（起初，也许哥白尼害怕被教会谴责为异端，所以将他的模型匿名地流传。）他的观念是，太阳静止地位于中心，而地球和行星们围绕着太阳作圆周运动。将近一个世纪以后，人们才认真接受他的观念。后来，两位天文学家——德国人约翰斯·开普勒和意大利人伽利略·伽利雷开始公开支持哥白尼理论，尽管它所预言的轨道还不能完全与观测相符合。直到1609年，亚里士多德和托勒密的理论才宣告死亡。那一年，伽利略用刚发明的望远镜来观测夜空。当他观测木星时，发现有几个小卫星或月亮围绕着它转动，这表明不像亚里士多德和托勒密设想的那样，并非所有东西都必须直接地围绕着地球转动（当然，仍然可能相信地球是静止地处于宇宙的中心，而木星的卫星沿着一种极其复杂的轨道围绕地球运动，表观上看来它们是围绕着木星转动。然而，哥白尼理论却简单得多了）。同时，约翰斯·开普勒修正了哥白尼理论，提出行星不是沿着圆周而是沿着椭圆（椭圆是拉长的圆）运动，从而最终使预言和观察相互一致了。

就开普勒而言，椭圆轨道仅仅是想当然的，并且是相当讨厌的假设，因为椭圆显然不如正圆那么完美。虽然他几乎偶然地发现椭圆轨道能很好地和观测相符合，但却不能把它和他的磁力引起行星围绕太阳运动的思想相互调和起来。只有到更晚得多的1687年，这一切才得到解释。这一年，艾萨克·牛顿爵士出版了他的《自然哲学的数学原理》，这部也许是物理科学中有史以来最重要的著作。在这部著作中，牛顿不但提出物体如何在空间和时间中运动的理论，并且发展了为分析这些运动所需的复杂的数学。此外，牛顿还提出了万有引力定律。根据这条定律，宇宙中的任一物体都被另外的物体吸引。物体质量越大，相互距离越近，则相互之间的吸引力越大。正是这同一种力，使物体下落到地面。（一个苹果落到牛顿的头上使他得到灵感的故事，几乎肯定是不足凭信的。牛顿自己说过的不过是，当他坐着陷入沉思之时，一个苹果的下落使他得到了万有引力的思想。）牛顿继而证明，根据他的定律，引力使月亮沿着椭圆轨道围绕着地球运行，而地球和其他行星沿着椭圆轨道围绕着太阳公转。

哥白尼的模型摆脱了托勒密的天球，以及与其相关的宇宙存在着自然边界的观念。“固定恒星”除了由于地球围绕着自身的轴自转引起的穿越天空的转动外，它们的位置显得固定不变，很自然会使人推测到固定恒星是和我们太阳类似的物体，只是比太阳离开我们远得多了。

按照他的引力理论，牛顿意识到恒星应该相互吸引，这样它们似乎不能保持基本上不动。难道它们不会都一起落到某处去吗？在1691年写给同时代另一位最重要的思想家理查德·本特里的一封信中，牛顿论证道，如果只有有限数目的恒星分布在一个有限的空间区域里，这确实是会发生的。但是另一方面，他推断说，如果存在无限数目的恒星，大体均匀地分布于无限的空间中，对它们而言，因为这时不存在一个中心落点，这种情形就不会发生。

当人们议论到无限时，这种论证是你会遭遇到的一种陷阱。在一个无限的宇宙中，因为在每一点的两边都有无限颗恒星，所以每一点都可以认为是中心。很久以后才意识到正确的方法，即是先考虑有限的情形，这时所有恒星都相互落到一起，然后加上在这个区域以外大体均匀分布的更多恒星，看事情会如何改变。按照牛顿定律，平均地讲，这额外的恒星对原先的那些根本没有什么影响，所以这些恒星还是同样快地落到一起。我们愿意加上多少恒星就可以加上多少，但是它们仍然总是向自身坍缩。现在我们知道，不可能存在一个无限静态的引力总是吸引的宇宙模型。

在20世纪之前从未有人提出过，宇宙是在膨胀或是在收缩，这有趣地反映了当时的思维风气。一般认为，宇宙要么以一种不变的状态存在了无限长的时间，要么以多多少少正如我们今天观察到的样子在有限久的过去创生。其部分的原因可能是，人们倾向于相信永恒的真理，也可能由于从以下的观念可以得到安慰，即虽然他们会生老病死，但是宇宙必须是不朽的不变的。

甚至那些意识到牛顿的引力理论导致宇宙不可能静止的人，也没有想到提出宇宙可能正在膨胀。相反，他们试图修正理论，使引力在非常大距离下变成排斥的。这并没有太大影响他们对行星运动的预言，然而却允许恒星的无限分布保持平衡状态——邻近恒星之间的吸引力被远距离外的恒星来的斥力平衡。然而，现在我们相信，这样的平衡是不稳定的：如果某一区域内的恒星稍微相互靠近一些，它们之间的引力就会增强，并超过斥力的作用，因此这些恒星就会继续落到一起。反之，如果某一区域内的恒星稍微相互远离一些，斥力就起主导作用，并驱使它们离得更远。

另一个反对无限静止宇宙的异见通常归功于德国哲学家亨利希·奥勃斯，他在1823年撰写了这个理论。事实上，牛顿的一些同时代人已经提出过这个问题。甚至奥勃斯的文章也不是貌似有理地反驳这个模型的第一篇。不管怎么说，这是第一篇被广泛注意到的文章。其困难在于，在一个无限静止的宇宙中，几乎每一道视线必须终结于某一颗恒星的表面。这样，人们可以预料，整个天空甚至在夜晚都会像太阳那么明亮。奥勃斯反驳说，远处恒星的光线会被它穿越过的物质吸收而减弱。然而如果真是如此，这介于其间的物质最终会被加热到发出和恒星一样强的光为止。可以避免整个天空像太阳那么明亮的结论的惟一方法是，假定恒星并非永远那么明亮，而是在有限久的过去才开始发光。在这种情况下，吸光物质还没加热，或者远处恒星的光线尚未到达我们这里。这就使我们面临着什么是首次引起恒星发光的问题。

当然，宇宙开端的问题比这早很久就被讨论过。根据一些早先的宇宙论和犹太教/基督教/穆斯林传统，宇宙在有限的并非非常遥远的过去的某个时刻启始。对于这样的一个开端，有一种论证是感到必须有“第一推动”来解释宇宙的存在。（在宇宙中，你总可以将一个事件解释为由另一个更早的事件引起的，但是只有当宇宙存在某个开端时，才能用这种方法解释它本身的存在。）圣·奥古斯丁在他的《上帝

之城》的著作中提出了另一种论证。他指出，文明在进步，我们将记住创造这些功绩或发展技术的人们。这样，人也许还有宇宙，不可能已经存在了那么长的时间。圣·奥古斯丁根据《创世纪》一书，接受公元前5000年作为宇宙创生的时刻。（有趣的是，这和最近一个冰河时代的结束，大约公元前10000年相距不远。考古学家告诉我们，文明实际上正是从那时开始的。）

另一方面，亚里士多德和其他大多数希腊哲学家不喜欢创生的思想，因为它带有太多的神学干涉的味道。所以他们相信，人类及其周围的世界已经并将继续永远存在。古人已经考虑到上述文明进步的论点，用周期性洪水或其他灾难的出现，使人类重复地回到文明的开初，来回答上面的诘难。

1781年，哲学家伊曼努尔·康德发表了里程碑般的（也是非常晦涩难懂的）著作——《纯粹理性批判》。在这本书中，他深入地考察了关于宇宙在时间上是否有开端、在空间上是否有限的问题。他称这些问题为纯粹理性的二律背反（也就是矛盾）。因为他感到存在同样令人信服的论据，来证明宇宙有开端的正命题，以及宇宙已经存在无限久的反命题。他对正命题的论证是：如果宇宙没有一个开端，则任何事件之前必有无限的时间。他认为这是荒谬的。他对反命题的论证是：如果宇宙有一开端，在它之前必有无限的时间，为何宇宙必须在某一特定的时刻开始呢？事实上，他对正命题和反命题用同样的论证来辩护。它们都是基于他隐含的假设，即不管宇宙是否存在了无限久，时间均可无限地倒溯回去。我们将会看到，在宇宙开端之前时间概念是没有意义的。这一点是圣·奥古斯丁首先指出的。当他被问及：“上帝在他创造宇宙之前做什么？”奥古斯丁并没有这样回答：“他正为诘问这类问题的人准备地狱。”而是说时间是上帝创造的宇宙的一个性质，在宇宙开端之前不存在。

当大部分人深信一个本质上静止不变的宇宙时，关于它有无开端的问题，实在是一个形而上学或神学的问题。按照宇宙存在无限久的

理论，或者按照宇宙以它似乎已经存在了无限久的样子在某一个有限时刻起始的理论，我们可以同样好地解释所观察到的事实。但在1929年，埃德温·哈勃作出了一个里程碑式的观测，即不管你往哪个方向观测，远处的星系都正急速地飞离我们而去。换言之，宇宙正在膨胀。这意味着，在早先的时刻星体更加相互靠近。事实上，似乎在大约100亿至200亿年之前的某一时刻，它们刚好在同一地方，所以那时候宇宙的密度为无限大。这个发现最终将宇宙开端的问题带进了科学的王国。

哈勃的发现暗示存在一个叫做大爆炸的时刻，当时宇宙的尺度无限小，而且无限紧密。在这种条件下，所有科学定律并因此所有预见将来的能力都崩溃了。如果在此时刻之前有过一些事件，它们将不可能影响现在发生的东西。因为它们没有任何观测的后果，所以可不理睬其存在。由于更早的时间根本没有定义，所以在这个意义上，人们可以说，时间在大爆炸时有一开端。必须强调的是，这个时间的开端和早先考虑的非常不同。在一个不变的宇宙中，时间的端点是必须由宇宙之外的存在物赋予的某种东西；宇宙的开端并没有物理的必然性。人们可以想像上帝在过去从字面上说的任何时刻创生了宇宙，如果宇宙正在膨胀，那么何以宇宙有一个开端似乎就有了物理的原因。人们仍然可以想像，上帝是在大爆炸的瞬间创生宇宙，或者甚至在更晚的时刻，以使它看起来就像发生过大爆炸似的方式创生，但是设想在大爆炸之前创生宇宙是没有意义的。大爆炸宇宙并没有排斥造物主，只不过对他何时从事这工作加上限制而已！

为了谈论宇宙的性质和讨论诸如它是否存在启始或终结的问题，你必须清楚什么是科学理论。我将采用素朴的观点，即理论只不过是宇宙或它的受限制的部分的模型，以及一套把这模型中的量和我们做的观测相联系的规则。它只存在于我们的头脑中，不再具有任何其他（不管在任何意义上）的实在性。一个好的理论必须满足以下两个要求：首先，这个理论必须能准确地描述大量的观测——这些观测是根

据只包含少数任选的元素的模型所做出的；其次，这个理论能对未来观测的结果作出明确的预言。例如，亚里士多德相信恩贝多克利的关于任何东西是由四元素即土、气、火和水组成的理论，该理论是足够简单的了，但它没有做出任何明确的预言。另一方面，牛顿的引力理论是基于甚至更为简单的模型，在此模型中两物体用一种力相互吸引，该力和被称为它们质量的量成正比，并和它们之间的距离的平方成反比。然而，它以很高的精确性预言了太阳、月亮和行星的运动。

在只是一个假设的意义上来讲，任何物理理论总是临时性的：你永远不可能证明它。不管多少回实验的结果和某个理论相一致，你永远不可能断定下一次结果不和它矛盾。另一方面，哪怕你只要找到一个和理论预言不一致的观测事实，即可证伪之。正如科学哲学家卡尔·波普强调的，一个好的理论的特征是，它能给出许多在原则上可以被观测否定或证伪的预言。每回观察到与这预言相符的新的实验，则这理论就存活，并且增加了我们对它的信任度；然而若有一个新的观测与之不符，则我们只得抛弃或修正这理论。

这被认为是迟早总会发生的事，但是你总可以质疑实现该观测的人员的能力。

在现实中经常发生的是，设计出的新理论实际上是原先理论的一个扩展。例如，非常精确地观测水星，发现它的运动和牛顿引力理论预言之间有一个微小的差异。爱因斯坦的广义相对论预言了和牛顿理论略微不同的运动。爱因斯坦的预言和观测到的相符合，而牛顿理论做不到，这个事实是对这个新理论的一个关键的证实。然而在我们正常处理的情形下，牛顿理论和广义相对论的预言之间差异非常小，所以为了所有实用的目的，我们仍然使用牛顿理论。（牛顿理论还有一个巨大的优点，用它计算比用爱因斯坦理论简单多了！）

科学的终极目的是提供描述整个宇宙的单一的理论。然而，大多数科学家遵循的方法是把问题分成两部分。首先，存在一些定律，这

些定律告诉我们宇宙如何随时间变化。（如果我们知道在任一时刻宇宙是什么样子的，这些定律就告诉我们它在未来任何时刻是什么样子的。）第二，存在宇宙初始状态的问题。有些人觉得科学只应关心第一部分，他们将初始状态的问题看作玄学或宗教的事体。他们会说，无所不能的上帝可以随心所欲地启始宇宙。那也许是真的，但是，倘若那样，他也可以使宇宙以完全任意的方式演化。可是，似乎他选择使宇宙以一种非常规则的，按照一定规律的方式演化。所以，看来可以同样合理地假定，也存在着制约初始状态的定律。

一蹴而就地设计一种能描述整个宇宙的理论，看来是非常困难的。相反，我们将这个问题分成许多小块，并发明许多部分理论。每一部分理论描述和预言一定有限范围的观测，同时忽略其他量的效应或用简单的一组数来代表。这方法可能全错。如果宇宙中的每一件东西都以非常基本的方式依赖于其他的任何一件东西，用隔离法研究问题的部分也许不可能逼近其完全的答案。尽管如此，这肯定是我们在过去取得进展的方法。牛顿引力理论又是一个经典的例子，它告诉我们两个物体之间的引力只取决于与每个物体相关的一个数——它的质量，而与物体由何物组成无关。这样，人们不需要太阳和行星结构和成分的理论就可以计算它们的轨道。

今天，科学家按照两个基本的部分理论——广义相对论和量子力学来描述宇宙。它们是本世纪上半叶的伟大的智慧成就。广义相对论描述引力和宇宙的大尺度结构，也就是从只有几英里直到大至1亿亿亿（1后面跟24个0）英里（1英里＝1.609千米），即可观测到的宇宙的尺度的结构。另一方面，量子力学处理极小尺度，例如万亿分之1英寸（1英寸＝2.54厘米）的现象。然而可惜的是，这两个理论不是相互协调的——它们不可能都对。当代物理学的一个主要的努力，以及本书的主题，即是寻求一个能将其合并在一起的新理论——量子引力论。我们还没有这样的理论，要获得这个理论，我们可能还有相当长的路要走，然而我们已经知道了这个理论所应具备的许多性质。在

以下几章，人们将会看到，我们对量子引力论所应有的预言已经知道得相当多。

现在，如果你相信宇宙不是任意的，而是被明确的定律制约的，你最终必须将这些部分理论合并成一个能描述宇宙中万物的完整统一理论。然而，在寻求这样的完整统一理论中有一个基本的矛盾。在前面概括的关于科学理论的思想中，假定我们是理性的生物，既可以随心所欲地观测宇宙，又可以从观察中得出逻辑推论。在这样的方案里可以合理地假设，我们可以越来越接近制约我们宇宙的定律。然而，如果真有一个完整的统一理论，则它大概也将决定我们的行动。这样，理论本身就决定了我们对之探索的结果！那么为什么它保证我们从证据得到正确的结论？难道它不也可以同样地保证我们引出错误的结论吗？或者根本没有结论？

对于这个问题，我所能给出的回答是基于达尔文的自然选择原理。该思想说，在任何自繁殖的群体中，总是存在着不同个体在遗传物质和发育上的变异。这些差异表明，某些个体比其他个体对它们周围的世界更能引出正确的结论，并去适应它。这些个体更可能存活、繁殖，因此它们的行为和思维的模式将越来越起主导作用。以下这一点在过去肯定是真的，即我们称之为智慧和科学发现的东西给我们带来了存活的好处。这种情况是否仍会如此没有这么清楚：我们的科学发现可以轻易地毁灭我们的一切。即使不是这样，一个完整的统一理论对于我们存活的机会不会有很大影响。然而，假定宇宙已经以规则的方式演化至今，我们可以预期，自然选择赋予我们的推理能力在探索完整统一理论时仍然有效，并因此不会导致我们得到错误的结论。

因为，除了最极端的情况外，我们已有的部分理论足以对所有的一切作出精确的预言，所以，要为探索宇宙的终极理论寻找实用的理由，看来就非常困难了。（值得指出，虽然类似的论点在过去既可以用来反对相对论。又可以用来反对量子力学，但这些理论已给我们带来了核能和微电子学的革命！）所以，发现一个完整的统一理论可能

对我们种族的存活无助，甚至也不会影响我们的生活方式。然而自从文明开始以来，人们即不甘心于将事件看作互不相关不可理解。他们渴望理解世界的根本秩序。今天我们仍然亟想知道，我们为何在此？我们从何而来？人类求知的最深切的意愿足以为我们从事的不断探索提供充足的理由。而我们的目标恰恰正是对于我们生存其中的宇宙作出完整的描述。

# 第二章 空间和时间

我们现在关于物体运动的观念来自于伽利略和牛顿。在他们之前，人们相信亚里士多德，他说物体的自然状态是静止的，并且只有在受到力或冲击的推动时才运动。这样，重的物体比轻的物体下落得更快，因为它受到更大的将其拉向地球的力。

亚里士多德的传统观点还以为，人们依靠纯粹思维即可以找出所有制约宇宙的定律：不必要用观测去检验之。这样，在伽利略之前，没有一个人想看看不同重量的物体是否确实以不同速度下落。据说，伽利略从比萨斜塔上将重物落下，从而证明了亚里士多德的信念是错的。这故事几乎不足以信，但是伽利略的确做了一些等效的事——让不同重量的球沿光滑的斜面上滚下。这情况类似于重物的垂直下落，只是因为速度小而更容易观察而已。伽利略的测量指出，不管物体的重量多少，其速度增加的速率是一样的。例如，你在一个沿水平方向每走10米即下降1米的斜面上释放1个球，则1秒钟后球的速度为每秒1米，2秒钟后为每秒2米，等等，而不管这个球多重。当然，一个铅锤比一片羽毛下落得更快些，那只是因为空气阻力将羽毛的速度降低。如果一个人释放两个不受任何空气阻力的物体，例如两个不同的铅锤，它们则以同样速度下降。在没有空气阻碍东西下落的月球上，航天员大卫，斯各特进行了羽毛和铅锤实验，并且发现两者确实同时落到月面上。

牛顿把伽利略的测量当做他的运动定律的基础。在伽利略的实验中，当物体从斜坡上滚下时，它一直受到不变外力（它的重量）的作

用，其效应是使它恒定地加速。这表明，力的真正效应总是改变物体的速度，而不是像原先想像的那样，仅仅使之运动。同时，它还意味着，只要物体没有受到外力，它就会以同样的速度保持直线运动。这一思想首次在牛顿于1687年出版的《数学原理》（即《自然哲学的数学原理》，下同——编者注）一书中明白地陈述出来，并被称为牛顿第一定律。牛顿第二定律给出物体在受力时发生的现象：物体在被加速或改变其速度时，其改变率与所受的外力成比例。（例如，如果力加倍，则加速度也将加倍。）物体的质量（或物质的量）越大，则加速度越小（以同样的力作用于具有两倍质量的物体时只产生一半的加速度）。小汽车可提供一个熟知的例子，发动机的功率越大，则加速度越大，但是小汽车越重，则对于同样的发动机，则加速度越小。除了他的运动定律，牛顿还发现了描述引力的定律：任何两个物体都相互吸引，其引力大小与每个物体的质量成比例。于是，如果其中一个物体（例如A）的质量加倍，则两个物体之间的引力加倍。这是你能预料得到的，因为新的物体A可看成两个具有原先质量的物体，每一个用原先的力来吸引物体B，所以A和B之间的总力加倍。而如果，比如说，其中一个物体质量大到原先的2倍，另一物体大到3倍，则引力就大到6倍。现在人们可以看到，为何落体总以同样的速率下降：具有两倍重量的物体受到将其向下拉的两倍的引力，但它的质量也大到两倍。按照牛顿第二定律，这两个效应刚好相互抵消，所以在所有情形下加速度都是相同的。

牛顿引力定律还告诉我们，物体之间的距离越远，则引力越小。牛顿引力定律讲，一个恒星的引力只是一个类似恒星在距离小一半时的引力的1/4。这个定律极其精确地预言了地球、月亮和其他行星的轨道。如果这定律中恒星的万有引力随距离减小或者增大得快一些，则行星轨道不再是椭圆的了，它们就会以螺旋线的形状要么盘旋到太阳上去，要么从太阳逃逸。

亚里士多德和伽利略-牛顿观念的巨大差别在于，亚里士多德相信

一个优越的静止状态，任何没有受到外力和冲击的物体都取这种状态。特别是他以为地球是静止的。但是从牛顿定律可以推断，并不存在惟一的静止标准。人们可以讲，物体A静止而物体B以不变的速度相对于物体A运动，或物体B静止而物体A运动，这两种讲法是等价的。例如，我们暂时不理睬地球的自转和它围绕太阳的公转，则可以讲地球是静止的，一辆有轨电车以每小时30英里的速度向东运动，或有轨电车是静止的，而地球以每小时30英里的速度向西运动。如果一个人在有轨电车上做运动物体的实验，所有牛顿定律仍然都成立。例如，在有轨电车上打乒乓球，人们将会发现，正如在铁轨旁一张台桌上的球一样，乒乓球服从牛顿定律，所以无法得知究竟是火车还是地球在运动。

缺乏静止的绝对标准意味着，人们不能确定，在不同时间发生的两个事件是否发生在空间的相同位置上。例如，假定在有轨电车上我们的乒乓球直上直下地弹跳，在1秒钟前后两次撞到桌面上的同一处。在铁轨上的人来看，这两次弹跳似乎发生在大约相距13米的不同的位置上，因为在这两回弹跳的时间间隔里，有轨电车已在铁轨上走了这么远。

这样，不存在绝对静止意味着不能像亚里士多德相信的那样，给事件指定一个绝对的空间位置。事件的位置以及它们之间的距离对于在有轨电车上和铁轨上的人来讲是不同的，所以没有理由以为一个人的立场比别人的更优越。

牛顿对不存在绝对位置或所谓绝对空间非常忧虑，因为这和他的绝对上帝的观念不一致。事实上，即使他的定律隐含着绝对空间的不存在，他也拒绝接受。因为这个非理性的信仰，他受到许多人的严厉批评，其中最有名的是贝克莱主教。他是一个相信所有的物质实体、空间和时间都是虚妄的哲学家。当人们将贝克莱的见解告诉著名的约翰逊博士时，他用脚趾踢到一块大石头上，并大叫道：“我要这样驳斥它！”

亚里士多德和牛顿都相信绝对时间。也就是说，他们相信人们可以毫不含糊地测量两个事件之间的时间间隔，只要用好的钟，不管谁去测量，这个时间都是一样的。时间相对于空间是完全分离并且独立的。这就是大部分人当作常识的观点。然而，我们必须改变这种关于空间和时间的观念。虽然这种显而易见的常识可以很好地对付运动甚慢的诸如苹果、行星的问题，但在处理以光速或接近光速运动的物体时却根本无效。

1676年，丹麦的天文学家欧尔·克里斯琴森·罗默第一次发现了，光以有限但非常高的速度旅行的事实。他观察到，木星的卫星不是以等时间间隔从木星背后出来，不像如果卫星以不变速度围绕木星运动时，人们会预料的那样。当地球和木星都围绕着太阳公转时，它们之间的距离在变化着。罗默注意到，我们离木星越远则木星的月食出现得越晚。他论证道，因为当我们离开更远时，光从木星卫星那里要花更长的时间才能达到我们这里。然而，他测得的木星到地球的距离变化不是非常准确，与现在的每秒186000英里的值相比较，那么他所测的光速的数值为每秒140000英里。尽管如此，罗默不仅证明了光以有限速度行进，并且测量了那个速度，他的成就是卓越的——要知道，这一切都是在牛顿发表《数学原理》之前1 1年做出的。

直到1865年，当英国的物理学家詹姆斯·麦克斯韦成功地将直到当时用以描述电力和磁力的部分理论统一起来以后，才有了光传播的正确的理论。麦克斯韦方程预言，在合并的电磁场中可以存在波动的微扰，它们以固定的速度，正如池塘水面上的涟漪那样行进。如果这些波的波长（两个相邻波峰之间的距离）为1米或更长一些，它们就是我们所谓的射电波。更短波长的波称做微波（几厘米）或红外线（长于万分之一厘米）。可见光的波长在一百万分之四十至一百万分之八十厘米之间。更短的波长被称为紫外线、X射线和伽马射线。

麦克斯韦理论预言，射电波或光波应以某一固定的速度行进。但是牛顿理论已经摆脱了绝对静止的观念，所以如果假定光以固定的速

度行进，人们就必须说清这固定的速度是相对于何物来测量的。因此有人提出，存在着一种无所不在的称为“以太”的物质，甚至在“真空的”空间中也是如此。正如声波在空气中行进一样，光波应该通过以太行进，所以它们的速度应是相对于以太而言的。相对于以太运动的不同观察者，会看到光以不同的速度冲他们而来，但是光对以太的速度保持不变。特别是当地球在它围绕太阳的轨道穿过以太时，在地球通过以太运动的方向测量的光速（当我们对光源运动时）应该大于在与运动垂直方向测量的光速（当我们不对光源运动时）。1887年，阿尔伯特·迈克耳孙（他后来成为美国第一位诺贝尔物理学奖获得者）和爱德华·莫雷在克里夫兰的凯思应用科学学校进行了一个非常仔细的实验。他们将沿地球运动方向以及垂直于此方向的光速进行比较。使他们大为惊奇的是，他们发现这两个光速完全一样！

在1887年至1905年之间，最著名者为荷兰物理学家亨得利克·洛伦兹做出的。然而，一位迄至当时还默默无名的瑞士专利局的职员阿尔伯特·爱因斯坦，在1905年的一篇著名的论文中指出，只要人们愿意抛弃绝对时间观念的话，整个以太的观念则是多余的。几个星期之后，法国第一流的数学家亨利·庞加莱也提出类似的观点。爱因斯坦的论证比庞加莱的论证更接近物理，后者将其考虑为数学问题。通常这个新理论归功于爱因斯坦，但人们不会忘记庞加莱的名字在其中起了重要的作用。

这个被称为相对论的基本假设是，不管观察者以任何速度作自由运动，相对于他们而言，科学定律都应该是一样的。这对于牛顿的运动定律当然是对的，但是现在这个观念被扩展到包括麦克斯韦理论和光速：不管观察者运动多快，他们应测量到一样的光速。这简单的观念有一些非凡的结论。可能最著名者莫过于质量和能量的等价，这可用爱因斯坦著名的方程$E=mc^2$来表达（E是能量，m是质量，c是光速），以及没有任何东西可能行进得比光还快的定律。由于能量和质量的等价，物体由于它的运动具有的能量应该加到它的质量上去。换

言之，要加速它将更为困难。这个效应只有当物体以接近于光速的速度运动时才有实际的意义。例如，以10%光速运动的物体的质量只比原先增加了0.5%，而以90%光速运动的物体，其质量变得比正常质量的2倍还多。当一个物体接近光速时，它的质量上升得越来越快，这样它需要越来越多的能量才能进一步加速上去。实际上它永远不可能达到光速，因为那时质量会变成无限大，而根据质量能量等价原理，这就需要无限大的能量才能做到。由于这个原因，相对论限制了物体运动的速度：任何正常的物体永远以低于光速的速度运动，只有光或其他没有内禀质量的波才能以光速运动。

相对论的一个同等非凡的推论是，它变革了我们空间和时间的观念。在牛顿理论中，如果有一光脉冲从一处发到另一处，（由于时间是绝对的）不同的观测者对这个行程所花的时间不会有异议，但是（因为空间不是绝对的）他们在光行进的距离上不会总取得一致的意见。由于光速正是它行进过的距离除以花费的时间，不同的观察者就测量到不同的光速。另一方面，在相对论中，所有的观察者必须在光以多快速度行进上取得一致意见。然而，在光行进过多远的距离上，他们仍然不能取得一致意见。因此，现在他们对光要花费多少时间上应该也不会取得一致意见。（花费的时间正是用光速——对这一点所有的观察者都意见一致——去除光行进过的距离——对这一点他们意见不一致。）换言之，相对论终结了绝对时间的观念！看来每个观察者都一定有他自己的时间测度，这是用他自己所携带的钟记录的，而不同观察者携带的同样的钟的读数不必要一致。

每个观察者都可以利用雷达发出光或射电波脉冲来说明一个事件在何处何时发生。一部分脉冲在事件反射回来后，观察者可在他接收到回波时测量时间。事件的时间可认为是脉冲被发出和反射被接收的两个时刻的中点：而事件的距离可取这来回行程时间的一半乘以光速（在这个意义上，一个事件是发生在空间的单独一点以及指定时间的一点的某件事）。这个思想被显示在（图2.1）上。这是时空图的一

个例子。利用这个步骤，作相互运动的观察者对同一事件可赋予不同的时间和位置。没有一个特别的观察者的测量比任何其他人的更正确，但是所有这些测量都是相关的。只要一个观察者知道其他人的相对速度，他就能准确算出其他人会赋予同一事件的时间和位置。

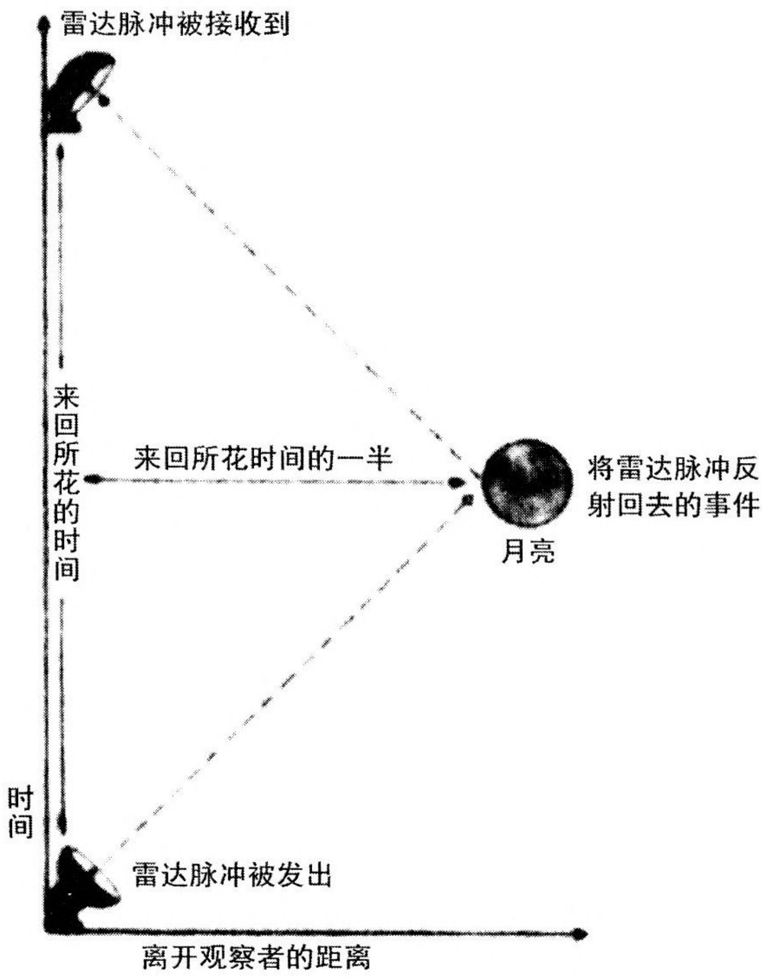



图2.1　时间用垂直坐标测量，离开观察者的距离用水平坐标测量。观察者在空间和时间里的路径用左边的垂线表示。到事件去和从事件来的光线的路径用对角线表示

现在我们正是用这种方法来准确地测量距离，因为我们可以把时

间比长度测量得更为准确。实际上，米是被定义为光在以铯原子钟测量的0.000000003335640952秒内行进的距离（取这个特别数字的原因是，因为它对应于历史上的米的定义——按照保存在巴黎的特定铂棒上的两个刻度之间的距离）。同样地，我们可以用叫做光秒的更方便的新长度单位，这就是简单地定义为光在1秒中行进的距离。现在，我们在相对论中按照时间和光速来定义距离，从而自然而然地，每个观察者都测量出光具有同样的速度（按照定义为每0.000000003335640952秒之一米）。没有必要引入以太的观念，正如迈克耳孙-莫雷实验显示的那样，以太的存在是无论如何检测不到的。然而，相对论迫使我们从根本上改变了我们的时间和空间观念。我们必须接受，时间不能完全脱离开和独立于空间，而必须和空间结合在一起形成所谓的时空的客体。

我们通常的经验是可以用3个数或坐标去描述空间中的一点的位置。譬如，人们可以说屋子里的一点离开一堵墙7英尺（1英尺＝0.3048米），离开另一堵墙3英尺，并且比地面高5英尺。或者人们也可以用一定的纬度、经度和海拔来指定该点。人们可以自由地选用任何3个合适的坐标，虽然它们只在有限的范围内有效。人们不是按照在伦敦皮卡迪里广场以北和以西多少英里以及高于海平面多少英尺来指明月亮的位置，取而代之，人们可用离开太阳、离开行星轨道面的距离以及月亮与太阳的连线和太阳与临近的一个恒星——例如半人马座α——连线之夹角来描述它的位置。甚至这些坐标对于描写太阳在我们星系中的位置，或我们星系在本星系群中的位置也没有太多用处。事实上，人们可按照一组相互交叠的坐标碎片来描写整个宇宙。在每一碎片中，人们可用不同的三个坐标的集合来指明点的位置。

一个事件是在特定时刻和在空间中特定的一点发生的某件事。这样，人们可以用4个数或坐标来指定它。再说一遍，坐标系的选择是任意的；人们可以使用任何3个定义好的空间坐标和任何时间测度。在相对论中，在时间和空间坐标之间没有真正的差别，犹如在任何两

个空间坐标之间没有真正的差别一样。人们可以选择一组新的坐标，比如说，第一个空间坐标是1日的第一和第二空间坐标的组合。例如，测量地球上一点的位置不用在伦敦皮卡迪里广场以北和以西的里数，而是用在它的东北和西北的里数。类似地，人们在相对论中可以用新的时间坐标，它是旧的时间（以秒作单位）加上往北离开皮卡迪里的距离（以光秒为单位）。

将一个事件的四坐标当作指定其在所谓的时空的四维空间中位置的手段经常是有助的。四维空间是不可想象的。对我个人来说，摹想三维空间已经足够困难！然而很容易画出二维空间图，例如地球的表面（地球的表面是两维的，因为可以用两个坐标，例如纬度和经度，来指明一点的位置。）我将通常使用二维图，向上增加的方向是时间，水平方向是其中一个空间坐标。不管另外两个空间坐标，或者有时用透视法将其中一个表示出来（这些被称为时空图，如图2.1所示）。例如，在图2.2中时间是向上的，并以年作为测量单位，而沿着从太阳到半人马座α连线的距离在水平方向上以英里为单位来测量。图中的左边和右边的垂线表示太阳和半人马座α通过时空的路径。从太阳发出的光线沿着对角线走，并且要花费4年的时间才能从太阳到达半人马座α。

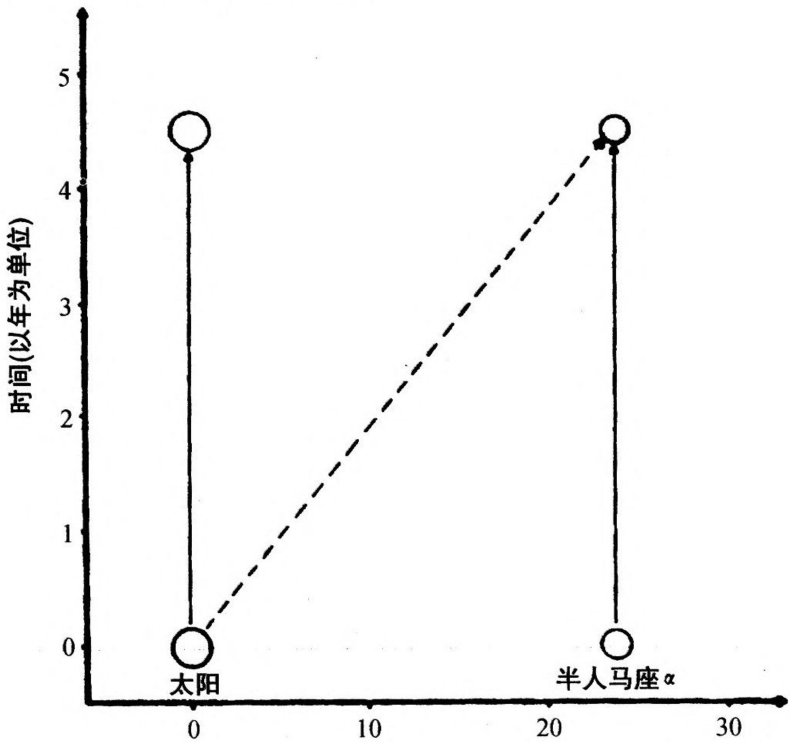



图2.2　离开太阳的距离[以$10^{12}$英里（1英里＝1.609千米）为单位]

正如我们已经看到的，麦克斯韦方程预言，不管光源的速度如何，光速应该是一样的，这已被精密的测量证实。由此推出，如果有一个光脉冲从一特定的空间点在一特定时刻发出，在时间的进程中，它就会作为一个光球面发散开来，而光球面的形状和大小与源的速度无关。在一百万分之一秒后，光就散开成一个半径为300米的球面；一百万分之二秒后，半径变成600米，等等。这正如同将一块石头扔到池塘里，水表面的涟漪向四周散开一样，涟漪作为一个圆周散开并随时间越变越大。如果人们把不同时刻涟漪的快照逐个堆叠起来，扩大的水波圆周就会画出一个圆锥，其顶点正是石块击到水面的地方和时刻。（图2.3）。类似地，从一个事件散开的光在（四维的）时空里形成了一个（三维的）圆锥，这个圆锥称为事件的将来光锥。以同样的方法可以画出另一个称为过去光锥的圆锥，它表示所有可以用一

个光脉冲传播到该事件的事件集合（图2.4）。

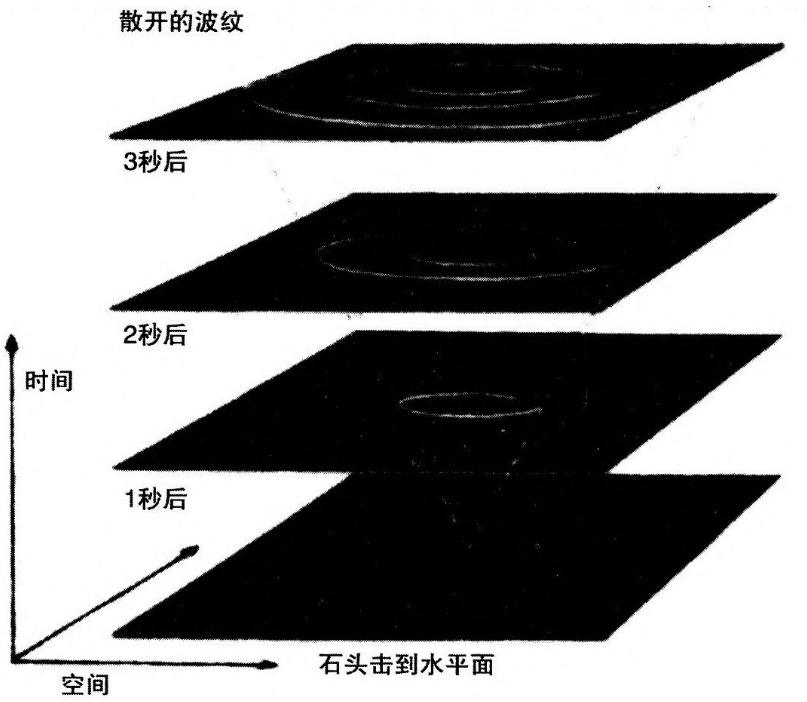



图2.3

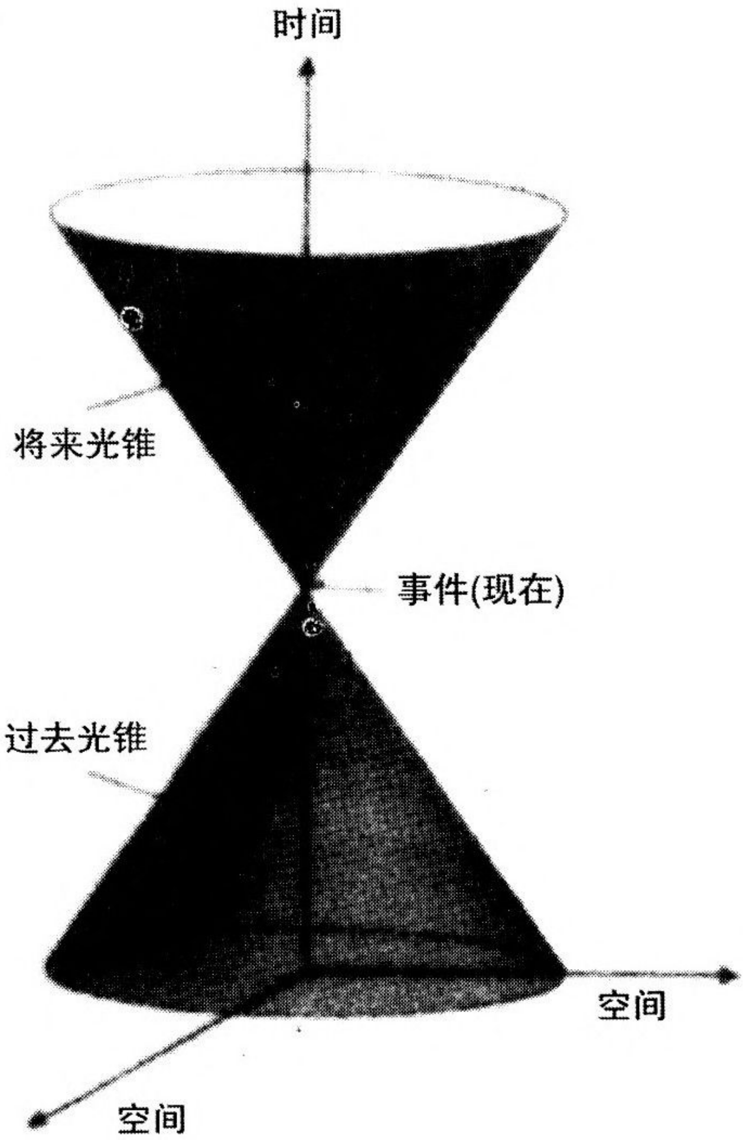



图2.4

对于给定的事件P，人们可以将宇宙中的其他事件分成三类。从事件P出发由一个粒子或者波以等于或小于光速的速度行进能到达的

那些事件称为属于P的将来。它们处于从事件P发射的膨胀的光球面之内或之上。这样，在时空图中它们就处于P的将来光锥的里面或上面。因为没有任何东西比光行进得更快，所以在P所发生的东西只能影响在P的将来中的事件。

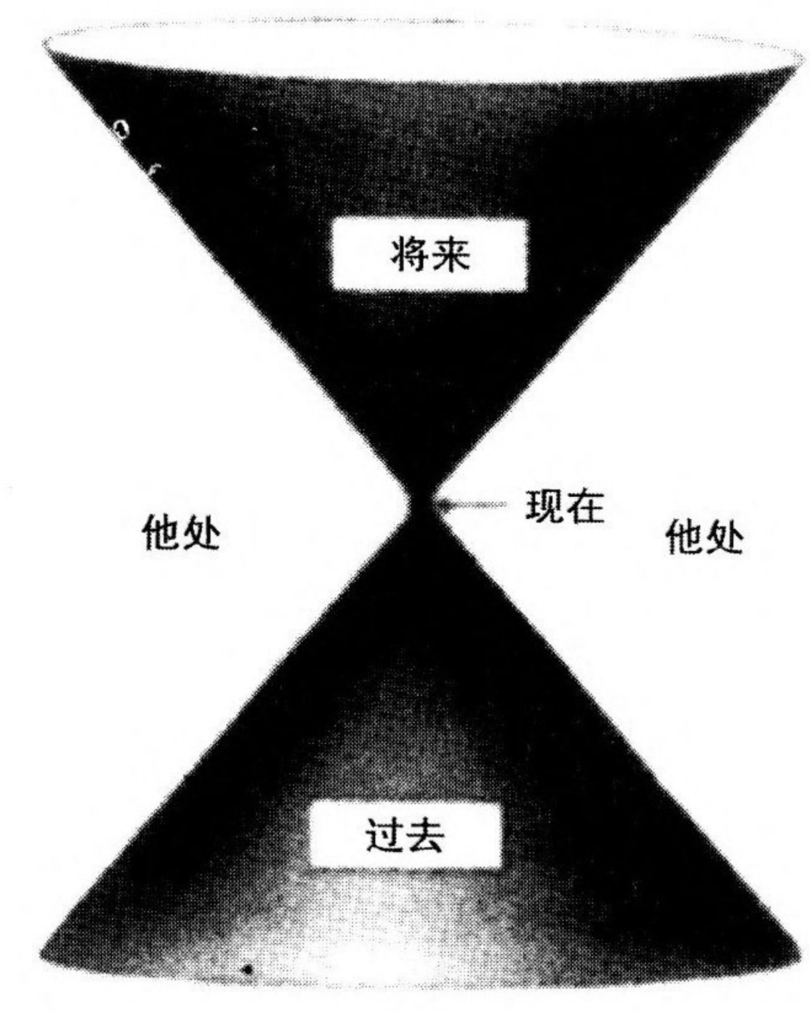



图2.5

类似地，P的过去可被定义为下述的所有事件的集合，从这些事

件可能以等于或小于光速的速度行进到达事件P。这样，它就是能够影响发生在P的事件的所有事件的集合。不处于P的将来或过去的事件被称之为处于P的他处（图2.5）。在这种事件处所发生的东西既不能影响发生在P的事件，也不受发生在P的事件的影响。例如，假定太阳就在此刻停止发光，它不会对此刻的地球上的事情发生影响，因为它们是在太阳熄灭这一事件的他处（图2.6）。我们只能在8分钟之后才知道这一事件，这是光从太阳到达我们所花费的时间。只有到那时候，地球上的事件才在太阳熄灭这一事件的将来光锥之内。类似地，我们也不知道这一时刻发生在宇宙中更远处的事：我们看到的从很远星系来的光是在几百万年之前发出的，至于我们看到的最远物体，光是在大约80亿年前发出的。这样，当我们看宇宙时，我们是在看它的过去。

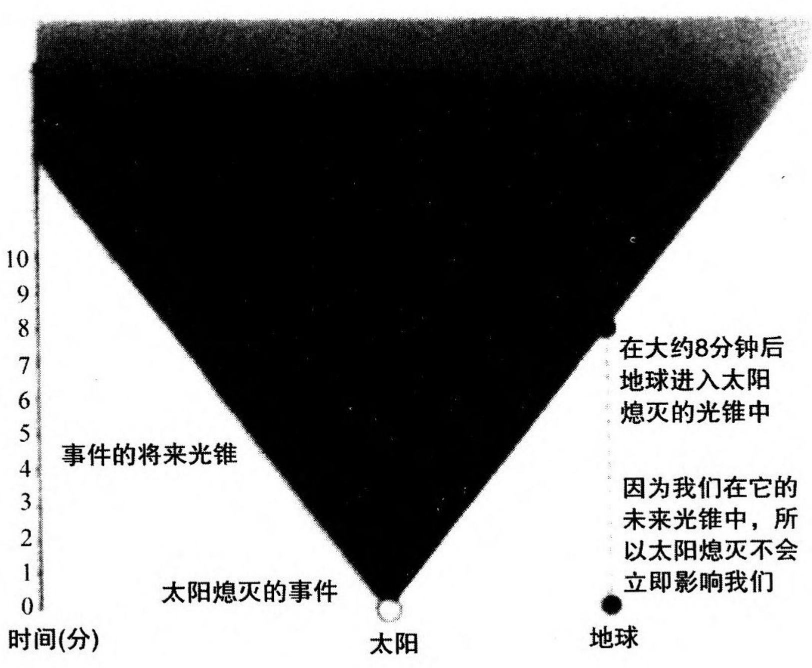



图2.6

如果人们忽略引力效应，正如爱因斯坦和庞加莱在1905年那样做的，人们就得到了称为狭义相对论的理论。对于时空中的每一事件我们都可以做一个光锥（所有从该事件发出的光的可能路径的集合），由于在每一事件处在任一方向上的光的速度都是一样的，所以所有光锥都是全等的，并朝着同一方向。这理论又告诉我们，没有任何东西行进得比光更快。这意味着，通过空间和时间的任何物体的轨迹必须由一根线来表示，而这根线落在它上面的每一事件的光锥之内（图2.7）。狭义相对论非常成功地解释了如下事实：对所有观察者而言，光速都是一样的（正如迈克耳孙——莫雷实验所展示的那样），并成功地描述了当物体以接近于光速运动时会发生什么。然而，它和牛顿引力理论不相协调。牛顿理论说，物体之间相互吸引，其吸引力依赖于它们之间的距离。这意味着，如果我们移动其中一个物体，另一物体所受的力就会立即改变。或换言之，引力效应必须以无限速度行进，而不像狭义相对论要求的那样，只能以等于或低于光速的速度行进。爱因斯坦在1908年至1914年之间进行了多次不成功的尝试，企图找到一个和狭义相对论协调的引力理论。1915年，他终于提出了今天我们称为广义相对论的理论。



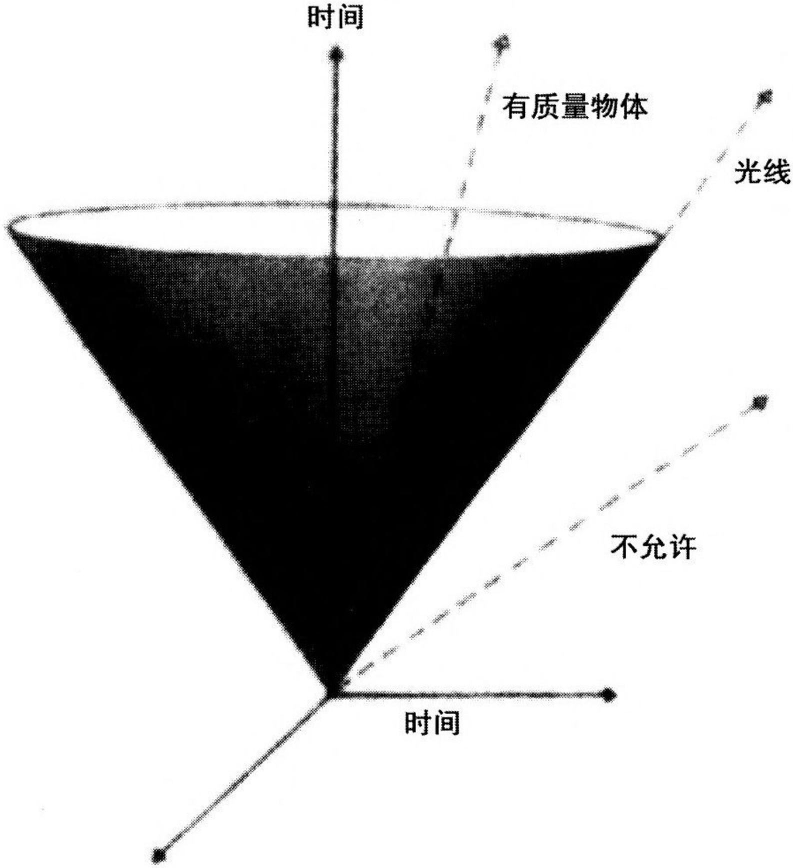



图2.7

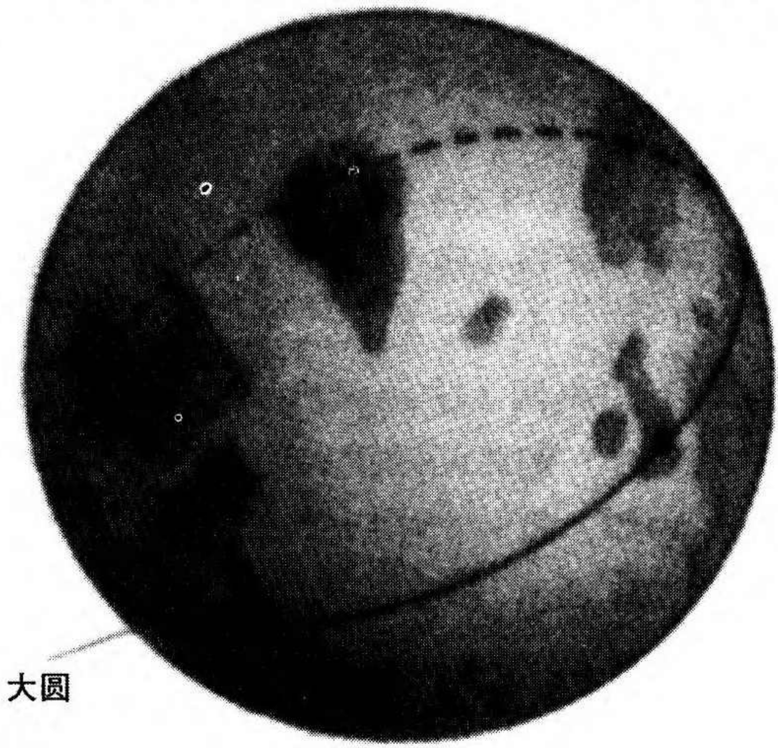



图2.8

爱因斯坦提出了革命性的思想，即引力不像其他种类的力，它只不过是时空不是平坦的这一事实的结果，而早先人们假定时空是平坦的。在时空中的质量和能量的分布使它弯曲或“翘曲”。像地球这样的物体并非由于称为引力的力使之沿着弯曲轨道运动，相反，它沿着弯曲空间中最接近于直线路径的东西运动，这个东西称为测地线。一根测地线是邻近两点之间最短（或最长）的路径。例如，地球的表面是个弯曲的二维空间。地球上的测地线称为大圆，是两点之间最近的路（图2.8）。由于测地线是两个机场之间的最短程，这正是领航员叫飞行员飞行的航线。在广义相对论中，物体总是沿着四维时空的直线走。尽管如此，在我们看来它在三维空间中是沿着弯曲的路径。（这正如同看一架在非常多山的地面上空飞行的飞机。虽然它沿着三维空间的直线飞，它在二维的地面上的影子却是沿着一条弯曲的路径。）

太阳的质量以这样的方式弯曲时空，使得在四维的时空中地球虽然沿着直线的路径，它却让我们看起来是沿着三维空间中的一个圆周轨道运动。事实上，广义相对论和牛顿引力理论预言的行星轨道几乎完全一致。然而，对于水星，这颗离太阳最近，受到引力效应最强，轨道被拉得相当长的行星，广义相对论预言其轨道椭圆的长轴应围绕着太阳以大约每1万年1度的速率进动。尽管这个效应如此微小，但在1915年前即被注意到了，并被作为爱因斯坦理论的第一个验证。近年来，其他行星和牛顿理论预言的甚至更小的轨道偏差已被雷达测量到，并且发现和广义相对论的预言相符。

光线也必须在时空中遵循测地线。时空是弯曲的事实再次意味着，光线在空间中看起来不是沿着直线行进。这样，广义相对论预言光线必须被引力场折弯。譬如，理论预言，由于太阳的质量的缘故，太阳近处的点的光锥会向内稍微弯折。这表明，从遥远恒星发出的刚好通过太阳附近的光线会被偏折很小的角度，对于地球上的观察者而言，这恒星似乎位于不同的位置（图2.9）。当然，如果从恒星来的光线总是在靠太阳很近的地方穿过，则我们就无从分辨，是光线被偏折了，还是该恒星实际上就在我们看到的地方。然而，由于地球围绕着太阳公转，不同的恒星显得从太阳后面通过，并且它们的光线受到偏折。所以，相对于其他恒星而言，它们改变了表观的位置。

在正常情况下，要观察到这个效应非常困难，这是由于太阳的光线使得人们不可能观看天空上出现在太阳附近的恒星。然而，在日食时就可能观察到，这时太阳的光线被月亮遮住了。由于第一次世界大战正在进行，爱因斯坦光偏折的预言不可能在1915年立即得到验证。直到1919年，一个英国的探险队从西非观测日食，证明光线确实像理论所预言的那样被太阳偏折。这次英国人证明德国人的理论被欢呼为战后两国和好的伟大行动。具有讽刺意味的是，后来人们检查这回探险所拍的照片，发现其误差和企图测量的效应同样大。他们的测量纯属运气，或是已知他们所要得的结果的情形，这在科学上时有发生。

然而，后来的多次观测准确地证实了光偏折。

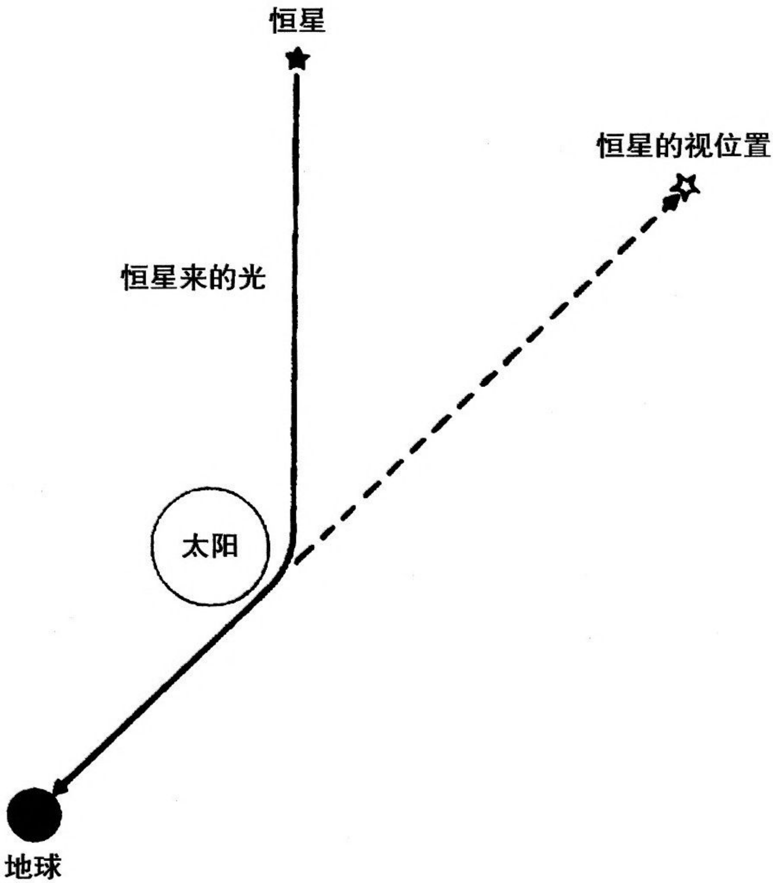



图2.9

广义相对论的另一个预言是，在像地球这样的大质量的物体附近，时间显得流逝得更慢一些。这是因为光能量和它的频率（光在每秒钟里波动的次数）有一种关系：能量越大，则频率越高。当光从地球的引力场往上行进，它失去能量，因而其频率下降（这表明两个相邻波峰之间的时间间隔变大）。在上面的某个人看来，下面发生的每一件事情都显得需要更长的时间。1962年，人们利用一对安装在水塔顶上和底下的非常准确的钟，验证了这个预言。发现底下的那只更接

近地球的钟走得较慢，这和广义相对论正好相符。目前，随着基于卫星信号的非常精确的导航系统的出现，地球上的不同高度的钟的速度的差异，在实用上具有相当的重要性。如果人们无视广义相对论的预言，计算的位置会错几英里。

牛顿运动定律使在空间中的绝对位置的观念寿终正寝。而相对论摆脱了绝对时间。考虑一对双生子。假定其中一个孩子去山顶上生活，而另一个留在海平面，第一个将比第二个老得快些。这样，如果他们再次相会，一个会比另一个更老一些。在这个例子中，年纪的差别会非常小。但是，如果有一个孩子在以近于光速运动的航天飞船中作长途旅行，这种差别就会大得多。当他回来时，他会比留在地球上另一个年轻得多。这叫做双生子佯谬，但是，只是对于头脑中仍有绝对时间观念的人而言，这才是佯谬。在相对论中并没有惟一的绝对时间，相反，每个人都有他自己的时间测度，这依赖于他在何处并如何运动。

1915年之前，空间和时间被认为是事件在其中发生的固定舞台，而它们不受在其中发生的事件的影响。即便在狭义相对论中，这也是对的。物体运动，力吸引并排斥，但时间和空间则完全不受影响地延伸着。空间和时间很自然地被认为无限地向前延伸。

然而在广义相对论中，情况则完全不同。这时，空间和时间变成为动力量：当物体运动，或者力作用时，它影响了空间和时间的曲率；反过来，时空的结构影响了物体运动和力作用的方式。空间和时间不仅去影响、而且被发生在宇宙中的每一件事影响。正如人们没有空间和时间的概念不能谈论宇宙的事件一样，同样地，在广义相对论中，在宇宙界限之外讲空间和时间也是没有意义的。

在以后的几十年中，对空间和时间的这种新理解是对我们宇宙观的变革。旧的宇宙观被新的宇宙观取代了。前者认为宇宙基本上是不变的，它可能已经存在了无限长的时间，并将永远继续存在下去；后

者则认为宇宙在运动、在膨胀，它似乎开始于过去的某一个时间，并也许会在将来的某一个时间终结。这个变革正是下一章的内容。几年之后，它又是我研究理论物理的起点。罗杰·彭罗斯和我证明了，爱因斯坦广义相对论意味着，宇宙必须有个开端，并且可能有个终结。

# 第三章 膨胀的宇宙

如果在一个清澈无月的夜晚仰望星空，人们能看到的最亮的星体最可能是金星、火星、木星和土星这几颗行星，还有巨大数目的正像我们太阳但离开我们远得多的恒星。事实上，随着地球围绕着太阳公转，某些固定的恒星相互之间的位置看起来确实起了非常微小的变化——它们不是完全固定不动的！这是因为它们距离我们较近一些。当地球围绕着太阳公转时，相对于更远处的恒星背景，我们从不同的位置观测它们。这是幸运的，因为它使我们能直接测量这些恒星离开我们的距离，它们离我们越近，就显得移动得越多。最近的恒星叫做比邻星，它离我们大约4光年那么远（从它发出的光大约花费4年才能到达地球），也就是大约23万亿英里的距离。其他大部分肉眼可见的恒星离开我们的距离均在几百光年之内。与之相比，太阳仅仅在8光分那么远！可见的恒星散布在整个夜空，但是特别集中在一条称为银河的带上。远在公元1750年，有些天文学家就提出，如果大部分可见的恒星处在一个单独的碟状的结构中，则银河的外观可以得到解释。这个结构便是今天我们称为螺旋星系的一个例子。之后不过几十年，天文学家威廉·赫歇尔爵士通过对大量恒星的位置和距离进行过细的编目分类，就证实了这个观念。即便如此，这个思想在本世纪初才完全被人们接受。

1924年，我们现代的宇宙图象才被奠定。那一年，美国天文学家埃德温·哈勃证明了，我们的星系不是惟一的星系。事实上，还存在其他许多星系，在它们之间是巨大的空虚的太空。为了证明这些，他必须确定这些星系的距离。这些星系是如此之遥远，不像邻近的恒星那

样，它们确实显得是固定不动的。所以哈勃被迫用间接的手段去测量这些距离。由于恒星的视亮度取决于两个因素：它辐射出来多少光（它的光度）以及它离我们多远。对于近处的恒星，我们可以测量其视亮度和距离，这样我们可以算出它的光度。相反，如果我们知道其他星系中恒星的光度，我们可用测量它们的视亮度来算出它们的距离。哈勃注意到，当某些类型的恒星近到足以被我们测量时，它们有相同的光度；所以他提出，如果我们在其他星系找出这样的恒星，我们可以假定它们有同样的光度——这样就可计算出那个星系的距离。如果我们能对同一星系中的许多恒星这样做，并且计算结果总是给出相同的距离，则我们就会相当地信赖自己的估计。

埃德温·哈勃用上述方法算出了9个不同星系的距离。现在我们知道，我们的星系只是用现代望远镜可以看到的几千亿个星系中的一个，每个星系本身都包含有几千亿颗恒星。图3.1所示的便是一个螺旋星系的图，从生活在其他星系中的人来看我们的星系，想必也类似这个样子。我们生活在一个宽约为10万光年并慢慢旋转着的星系中；在它的螺旋臂上的恒星围绕着它的中心公转一圈大约花费几亿年。我们的太阳只不过是一颗平常的、平均大小的、黄色的恒星，它位于一个螺旋臂的内边缘附近。我们离开亚里士多德和托勒密的观念肯定相当远了，那时人们认为地球是宇宙的中心！

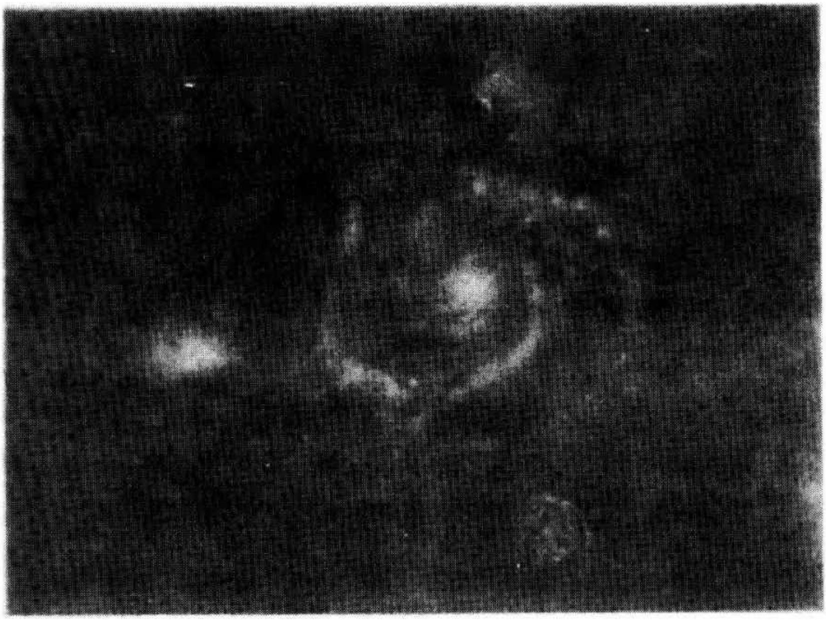


图3.1

恒星离开我们是如此之遥远，使我们只能看到极小的光点，而看不到它们的大小和形状。这样怎么能区分不同的恒星种类呢？对于绝大多数的恒星而言，只有一个特征可供观测——光的颜色。牛顿发现，如果太阳光通过一个称为棱镜的三角形状的玻璃块，就会被分解成像在彩虹中一样的分颜色（它的光谱）。将一台望远镜聚焦在一个单独的恒星或星系上，人们就可类似地观察到从这恒星或星系来的光谱。不同的恒星具有不同的光谱，但是不同颜色的相对亮度总是和人们期望从一个红热的物体发出的光的光谱完全一致。（实际上，从任何不透明的灼热的物体发出的光，有一个只依赖于它的温度的特征光谱——热谱。这意味着可以从恒星的光谱得知它的温度。）此外，我们发现，某些非常特定的颜色在恒星光谱里丢失，这些失去的颜色可依不同的恒星而异。由于我们知道，每一化学元素吸收非常独特的颜色族系，将它们和恒星光谱中失去的颜色相比较，我们就可以准确地确定恒星大气中存在哪种元素。

在20年代，当天文学家开始观察其他星系中的恒星光谱时，他们发现了某些最奇异的现象：它们和我们的银河系一样具有吸收的特征线族，只是所有这些线族都向光谱的红端移动了同样的相对量。为了理解其含意，我们必须首先理解多普勒效应。正如我们已经看到的，可见光由电磁场的起伏或波动构成。光的波长（或者相邻波峰之间的距离）极其微小，约为0.0000004至0.0000008米。光的不同波长正是人眼看成不同颜色的东西，最长的波长出现在光谱的红端，而最短的波长在光谱的蓝端。现在想像在离开我们固定的距离处有一个光源——例如一颗恒星——以固定的波长发出光波。显然，我们接收到的波长和发射时的波长一样（星系的引力场没有强到足以对它产生明显的效应）。现在假定这恒星光源开始向我们运动。当光源发出第二个波峰时，它离开我们较近一些，这样两个波峰之间的距离比恒星静止时较小。这意味着，我们接收到的波的波长比恒星静止时较短。相应地，如果光源离开我们运动，我们接收的波的波长将较长。这意味着，当恒星离开我们而去时，它们的光谱向红端移动（红移），而当恒星趋近我们而来时，光谱则被蓝移。这个称作多普勒效应的频率和速度的关系是我们日常熟悉的。例如听一辆小汽车在路上驶过：当它趋近时，它的发动机的音调变高（对应于声波的短波长和高频率）；当它经过我们身边而离开时，它的音调变低。光波或射电波的行为与之类似。警察就是利用多普勒效应的原理，靠测量射电波脉冲从车上反射回来的波长来测定车速。

在哈勃证明了其他星系存在之后的几年里，他花时间为它们的距离编目以及观察它们的光谱。那时候大部分人都以为，这些星系完全随机运动，所以预料会发现和红移光谱一样多的蓝移光谱。因此，当他发现大部分星系是红移的：几乎所有都远离我们而去时，确实令人十分惊异！1929年哈勃发表的结果更令人惊异：甚至星系红移的大小也不是随机的，而是和星系离开我们的距离成正比。或换句话讲，星系越远，它离开我们运动得越快！这表明宇宙不能像人们原先所想像的那样处于静态，而实际上是在膨胀；不同星系之间的距离一直在增

加着。

宇宙膨胀的发现是20世纪最伟大的智力革命之一。事后想起来，何以过去从来没有人想到这一点？！牛顿或其他人早就应该意识到，静态的宇宙在引力的影响下会很快开始收缩。然而现在假定宇宙正在膨胀，如果它膨胀得相当慢，引力就会使之最终停止膨胀，然后开始收缩。但是，如果它以比某一临界率更大的速度膨胀，引力则永远不足够强到使它停止膨胀，宇宙就永远继续膨胀下去。这有点像当一个人在地球表面引燃火箭上天时发生的情形，如果火箭的速度相当小，引力将最终使火箭停止并折回地面；另一方面，如果火箭具有比某一临界值（大约每秒7英里）更大的速度，引力的强度就不足以将其拉回，这样它将继续永远飞离地球。19世纪、18世纪甚至17世纪晚期的任何时候，人们都可以从牛顿的引力论预言出宇宙的这个行为。然而，静态宇宙的信念是如此之强，以至于一直维持到了20世纪的早期。甚至爱因斯坦于1915年发表其广义相对论时，还是这么肯定宇宙必须是静态的，以至于他在其方程中引进一个所谓的宇宙常数来修正自己的理论，使静态的宇宙成为可能。爱因斯坦引入一个新的“反引力”，这力不像其他力那样，不由任何特别的源引起，而恰恰是时空结构固有的。他宣称，时空有一内在的膨胀的趋向，这可以用来刚好去平衡宇宙间所有物质的相互吸引，由此导致一个静态的宇宙。当爱因斯坦和其他物理学家正在想方设法避免广义相对论的非静态宇宙的预言时，看来只有一个人，即俄国物理学家和数学家亚历山大·弗里德曼愿意只用广义相对论着手解释它。

弗里德曼对于宇宙作了两个非常简单的假定：我们不论往哪个方向看，也不论在任何地方进行观察，宇宙看起来都是一样的。弗里德曼指出，仅仅从这两个观念出发，我们就应该预料宇宙不是静态的。事实上，弗里德曼在1922年所做的预言，正是几年之后埃德温·哈勃观察到的结果。

很清楚，关于宇宙在任何方向上都显得一样的假设，实际上是不

对的。例如，正如我们看到的，我们星系中的其他恒星形成了横贯夜空的叫做银河系的光带。但是如果看得更远，星系数目则或多或少显得是相同的。所以假定我们在比星系间距离更大的尺度下来观察，而不管在小尺度下的差异，则宇宙确实在所有的方向看起来是大致一样的。在很长的时间里，这为弗里德曼的假设——作为实际宇宙的粗糙近似提供了充分的理由。但是，近世出现的一桩幸运事件揭示了以下事实，弗里德曼假设实际上异常准确地描述了我们的宇宙。

1965年，美国新泽西州贝尔电话实验室的两位美国物理学家阿诺·彭齐亚斯和罗伯特·威尔逊正在检测一个非常灵敏的微波探测器。（微波正如光波，但是它的波长大约为1厘米。）他们的探测器收到了比预想的还要大的噪声。彭齐亚斯和威尔逊为此而忧虑，这噪声不像是从任何特别的方向来的。首先他们在探测器上发现了鸟粪并检查了其他可能的故障，但很快就排除了这些可能性。他们知道，当探测器倾斜地指向天空时，从大气层里来的任何噪声都应该比原先垂直指向时更强，因为从接近地平线方向接收比起直接从头顶方向接收，光线要穿过多得多的大气。然而，不管探测器朝什么方向，这额外的噪声都是一样的，所以它一定是从大气层以外来的。并且，它在白天、夜晚、整年都是一样，尽管地球围绕着自己的轴自转或围绕太阳公转。这表明，这辐射必须来自太阳系以外，甚至星系之外，否则，当地球的运动使探测器指向不同方向时，噪声就会变化。

事实上，我们知道这辐射必须穿过我们可观察到的宇宙的大部分才行进至此，并且由于它在不同方向上都一样，如果只在大尺度下，这宇宙也必须是各向同性的。现在我们知道，不管我们朝什么方向看，这噪声的变化总是非常微小：这样，彭齐亚斯和威尔逊无意中非常精确地证实了弗里德曼的第一个假设。然而，由于宇宙并非在每一个方向上，而是在大尺度的平均上完全相同，所以微波也不可能在每一个方向上完全相同。在不同的方向之间必须有一些小变化。1992年宇宙背景探险者，或称为COBE，首次把它们检测到，其幅度大约为

十万分之一。尽管这些变化很小，但是正如我们将在第八章解释的，它们非常重要。

大约与彭齐亚斯和威尔逊在研究探测器中的噪声的同时，在附近的普林斯顿大学的两位美国物理学家，罗伯特·狄克和詹姆斯·皮帕尔斯也对微波感兴趣。他们正在研究乔治·伽莫夫（曾为亚历山大·弗里德曼的学生）的一个见解：早期的宇宙一定是非常密集的白热的。狄克和皮帕尔斯认为，我们应该仍然能看到早期宇宙的白热，这是因为从它的非常远的部分来的光，刚好现在才到达我们这里。然而，宇宙的膨胀把光红移得如此厉害，现在只能作为微波辐射被我们观察到。正当狄克和皮帕尔斯准备寻找这辐射时，彭齐亚斯和威尔逊听到了他们的工作，并且意识到，他们自己已经找到了它。为此，彭齐亚斯和威尔逊被授予1978年的诺贝尔奖（狄克和皮帕尔斯看来有点难过，更别提伽莫夫了）。

现在初看起来，关于宇宙在任何方向看起来都一样的所有证据似乎暗示，我们在宇宙中的位置有点特殊。特别是，如果我们看到所有其他的星系都远离我们而去，那似乎我们必须在宇宙的中心。然而，还存在另外的解释：从任何其他星系上看宇宙，在任何方向上也都一样。正如我们已经看到的，这是弗里德曼的第二个假设。我们没有任何科学的证据去相信或反驳这个假设。我们之所以相信它只是基于谦虚：因为如果宇宙只在围绕我们的所有方向显得相同，而在围绕宇宙的其他点却并非如此，则是非常令人惊奇的！在弗里德曼模型中，所有的星系都相互直接离开。这种情形很像一个画上好多斑点的气球被逐渐吹胀。当气球膨胀时，任何两个斑点之间的距离加大，但是没有一个斑点可认为是膨胀的中心。此外，斑点相离得越远，则它们相互离开得越快。类似地，在弗里德曼的模型中，任何两个星系相互离开的速度和它们之间的距离成正比。所以人们预言，星系的红移应与离开我们的距离成正比，这正是哈勃发现的。尽管他的模型取得了成功并预言了哈勃的观测，但是直到1935年，为了响应哈勃的宇宙均匀膨

胀的发现，美国物理学家霍瓦德·罗伯逊和英国数学家阿瑟·瓦尔克发现了类似的模型后，弗里德曼的工作才在西方被普遍知道。

虽然弗里德曼只找到一个模型，其实满足他的两个基本假设的共有三类模型。在第一类模型（即弗里德曼找到的）中，宇宙膨胀得足够慢，这样不同星系之间的引力使膨胀减缓，并最终停止。然后星系开始相互靠近，而宇宙收缩。图3.2表示随着时间增加两个邻近星系之间距离的变化。刚开始时距离为零，接着它增长到最大值，然后又减小到零；在第二类解中，宇宙膨胀得如此之快，引力虽然能使之缓慢一些，却永远不能使之停止。图3.3展示在此模型中的邻近星系之间的距离。刚开始时距离为零，最后星系以稳恒的速度相互离开；最后，还有第三类解，宇宙的膨胀快到足以刚好避免坍缩。正如图3.4所示的，星系的距离也从零开始，然后永远增大。然而，虽然星系分开的速度永远不会完全变为零，但是却会越变越小。

第一类弗里德曼模型的奇异特点是，宇宙在空间上不是无限的，但却没有边界。引力如此强大，将空间折弯使之再绕回到自身，这样就和地球的表面相当类似。如果有人在地球的表面上沿着一定的方向不停地旅行，他将永远不会遇到一个不可超越的障碍或从边缘掉下去，反而最终回到他出发的那一点。第一类弗里德曼模型中的空间与此非常相像，只不过地球表面是二维的，而它是三维的罢了。第四维时间在范围上也是有限的，然而它像一根有两个端点或边界即开端和终端的线。以后我们会看到，当人们将广义相对论和量子力学的不确定性原理结合在一起时，就可能使空间和时间都成为有限的，而没有任何边缘或边界。

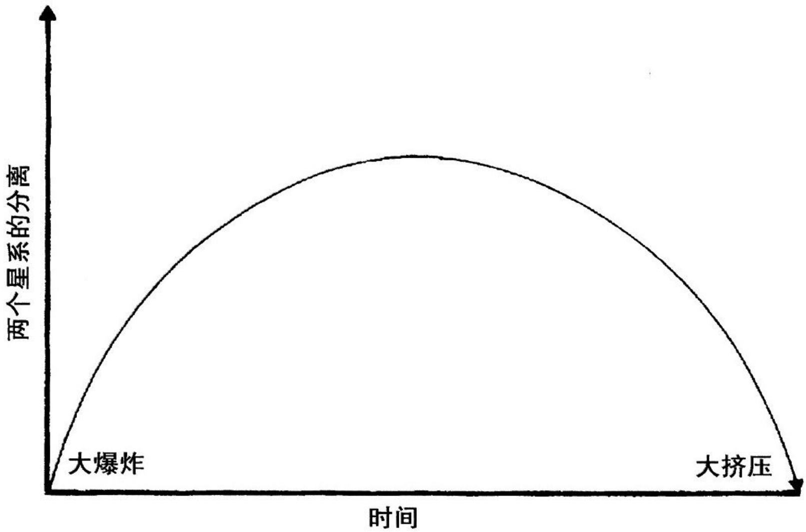



图3.2

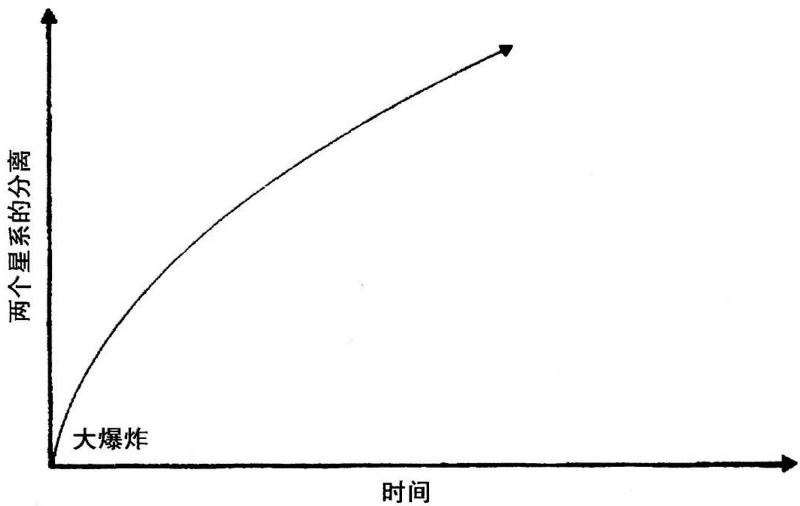



图3.3

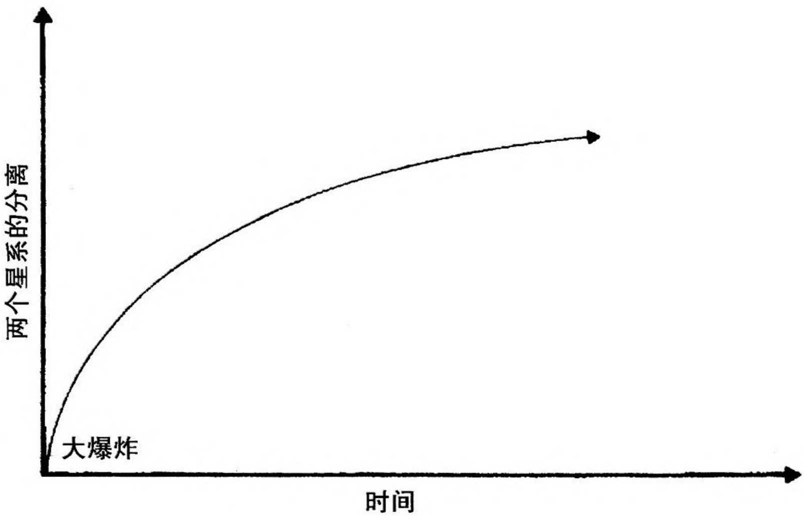



图3.4

一个人可以绕宇宙一周最终回到出发点的思想是科学幻想的好题材，但它在实际上并没有多大意义。因为可以证明，一个人还没来得及绕回一圈，宇宙已经坍缩到了零尺度。你必须旅行得比光还快，才能在宇宙终结之前绕回到你的出发点——而这是不允许的！

在第一类弗里德曼模型中，宇宙膨胀后又坍缩，空间如同地球表面那样，弯曲后又折回到自身。在第二类永远膨胀的模型中，空间以另外的方式弯曲，如同一个马鞍面。所以，在这种情形下，空间是无限的。最后，在第三类刚好以临界速率膨胀的弗里德曼模型中，空间是平坦的（而因此也是无限的）。

但是究竟何种弗里德曼模型描述我们的宇宙呢？宇宙最终会停止膨胀并开始收缩，还是将永远膨胀下去吗？要回答这个问题，我们必须知道现在的宇宙膨胀速度和它现在的平均密度。如果密度比一个由膨胀率决定的临界值还小，则引力太弱不足以将膨胀停止；如果密度比这临界值大，则引力会在未来的某一时刻将膨胀停止并使宇宙坍缩。

利用多普勒效应，可由测量星系离开我们的速度来确定现在的膨胀速度。这可以非常精确地实现。然而，因为我们只能间接地测量星系的距离，所以它们的距离知道得不很清楚。我们知道的不过是，宇宙在每10亿年里膨胀5%～10%。然而，我们对现在宇宙的平均密度测量得更不准确。我们如果将银河系和其他星系的所有能看到恒星的质量加起来，甚至按对膨胀率的最低的估值而言，其质量总量还不到用以阻止膨胀的临界值的1%。然而，在我们以及其他星系里应该包含大量的“暗物质”，那是我们不能直接看到的，但由于它的引力对星系中恒星轨道的影响，我们知道它必定存在。此外人们发现，大多数星系是成团的。我们能类似地推断，由其对星系运动的效应，在这些成团的星系之间还存在更多的暗物质。将所有这些暗物质加在一起，我们仍只能获得为停止膨胀必需的密度的1/10左右。然而，我们不能排除这样的可能性，可能还有我们尚未探测到的其他的物质形式，它们几乎均匀地分布于整个宇宙中，它仍可能使得宇宙的平均密度达到停止膨胀所必需的临界值。所以，现在的证据暗示，宇宙可能会永远地膨胀下去。但是，所有我们能真正肯定的是，既然它已经至少膨胀了100亿年，即便宇宙将要坍缩，至少要再过这么久才有可能。这不应使我们过度忧虑——到那时候，除非我们已到太阳系以外开拓了殖民地，否则人类早就随着太阳的消灭而死亡殆尽！

所有的弗里德曼解都具有一个特点，即在过去的某一时刻（约100至200亿年之前）邻近星系之间的距离一定为零。在这被我们称之为大爆炸的那一时刻，宇宙的密度和时空曲率都是无限大。因为数学不能真正地处理无限大的数，这意味着，广义相对论（弗里德曼解以此为基础）预言，在宇宙中存在一点，在该处理论本身崩溃。这样的点正是数学中称为奇点的一个例子。事实上，我们所有的科学理论都是基于时空是光滑的和几乎平坦的基础上表述的，所以它们在时空曲率为无限大的大爆炸奇点处崩溃。这意味着，即使在大爆炸前存在事件，人们也不能用它们去确定其后所要发生的事件，因为可预见性在大爆炸处崩溃了。

相应地，如果，事实也正是如此，我们只知道在大爆炸后发生的事件，我们就不能确定在这之前发生什么。就我们而言，发生于大爆炸之前的事件不能有后果，所以并不构成我们宇宙的科学模型的一部分。因此，我们应将它们从模型中割除掉，并宣称时间是从大爆炸开始的。

很多人不喜欢时间有个开端的观念，可能是因为它略带有神的干涉的味道。（另一方面，天主教会抓住了大爆炸模型，并在1951年正式宣告，它和《圣经》相和谐。）因此，人们多次企图避免曾经存在过大爆炸的这一结论。所谓的稳态理论得到过最广泛的支持。这是由纳粹占领的奥地利来的两个难民——赫曼·邦迪和托马斯·高尔德，以及一个在战时和他们一道从事雷达研制的英国人，弗雷德·霍伊尔于1948年共同提出的。其想法是，当星系相互离开时，由正在连续产生的新物质在它们中的间隙不断地形成新的星系。因此，在空间的所有点以及在所有的时间，宇宙看起来在大致上是相同的。稳态理论需要对广义相对论进行修正，使之允许物质的连续生成，但是有关的产生率是如此之低（大约每年每立方千米一个粒子），低到不与实验相冲突。在第一章叙述的意义上，这是一个好的科学理论：它非常简单，并做出确定的预言可让观察者检验。其中一个预言是，我们无论在宇宙的何时何地看给定的空间体积内星系或类似物体的数目必须一样。20世纪50年代晚期和60年代早期，由马丁·赖尔（他在战时也和邦迪、高尔德以及霍伊尔共事，作雷达研究）领导的一个天文学家小组在剑桥对从外空间来的射电源进行了普查。这个剑桥小组指出，这些射电源的大多数必须位于我们星系之外（它们中的许多确实可被认证与其他星系相关），并且存在的弱源比强源多得多。他们将弱源解释为较远的源，强源为较近的源。结果发现，单位空间体积内普通的源似乎在近处比远处稀少。这可能表明，我们处于宇宙的一个巨大区域的中心，这里的源比其他地方稀少。另外的一个解释是，宇宙在射电波向我们出发的过去的那一时刻具有比现在更密集的源。任何一种解释都和稳态理论相矛盾。此外，1965年彭齐亚斯和威尔逊的微波背景

辐射的发现还指出，宇宙在过去必定密集得多。因此必须抛弃稳态理论。

1963年，两位苏联科学家欧格尼·利弗席兹和艾萨克·哈拉尼可夫做了另外的尝试，设法避免存在大爆炸并因此引起时间起点的问题。他们提出，大爆炸可能只是弗里德曼模型的特性，这个模型毕竟只是真实宇宙的近似。也许，在所有大体类似实在宇宙的模型中，只有弗里德曼模型包含大爆炸奇点。在弗里德曼模型中，所有星系都直接相互离开——所以一点都不奇怪，在过去的某一时刻它们必须在同一处。然而，在实际的宇宙中，星系不仅仅直接相互离开——它们还有一些斜向速度。所以，在实际上它们从来没必要恰好在同一处，只不过非常靠近而已。也许，现在膨胀着的宇宙不是来自于大爆炸奇点，而是来自于更早期的收缩相；当宇宙坍缩时，其中的粒子可以不都碰撞，而是相互离得很近飞过然后又离开，产生了现在的宇宙膨胀。那么何以得知这实际的宇宙是否从大爆炸起始的呢？利弗席兹和哈拉尼可夫所做的，是去研究大体和弗里德曼模型相像的宇宙模型，但是考虑了实际宇宙中的星系的不规则性和杂乱速度。他们指出，即使星系不再总是直接相互离开，这样的模型也可以从一个大爆炸开始。但是他们宣称，这只在某些例外的模型中仍然可能发生，那里所有星系都以正确的方式运动。他们论证道，似乎没有大爆炸奇点的类弗里德曼模型比有此奇点的模型多无限多倍，所以我们的结论应该是，在实际上并没有过大爆炸。然而，他们后来意识到，存在更为广泛的具有奇点的类弗里德曼模型，那里的星系不必以任何特别的方式运动。所以，他们在1970年撤回了自己的主张。

利弗席兹和哈拉尼科夫的工作是有价值的。因为它显示了，如果广义相对论是正确的，宇宙可以有过奇点，一个大爆炸。然而，它没有解决关键的问题：广义相对论是否预言我们的宇宙一定有过大爆炸或时间的开端？对于这个问题，英国数学家兼物理学家罗杰·彭罗斯在1965年以完全不同的手段给出了回答。利用广义相对论中光锥行为的

方式以及引力总是吸引这个事实，他证明了，坍缩的恒星在自己的引力作用下陷入到一个区域之中，其表面最终缩小到零。并且由于这区域的表面缩小到零，它的体积也应如此。恒星中的所有物质将被压缩到一个零体积的区域里，所以物质的密度和时空的曲率变成无限大。换言之，人们得到了一个奇点，它被包含在一个叫做黑洞的时空区域中。

彭罗斯的结果乍看起来只适用于恒星，它并没有涉及任何关于整个宇宙的过去是否有个大爆炸奇点的问题。然而，当彭罗斯在创作他的定理之时，我还是一名研究生，正在尽力寻求一个完成博士论文的问题。两年之前我即被诊断得了肌萎缩性（脊椎）侧索硬化症，通常又称为卢伽雷病或运动神经细胞病，并且得知只有一两年可活了。在这种情况下，看来没有很多必要攻读博士学位了——我预料不能活那么久。然而两年过去了，我没有糟糕到那种程度。事实上，我的事情还进行得相当好，还和一个非常好的姑娘简·瓦尔德订婚了。但是为了结婚，我需要一份工作；为了得到工作，我需要一个博士学位。

1965年，我读到彭罗斯关于任何物体受到引力坍缩必定最终形成一个奇点的定理。我很快意识到，如果人们将彭罗斯定理中的时间方向颠倒以使坍缩变成膨胀，假定现在宇宙在大尺度上大体类似弗里德曼模型，这定理的条件仍然成立。彭罗斯定理已经指出，任何坍缩星必定终结于一个奇点；其时间颠倒的论证则是，任何类弗里德曼膨胀宇宙一定是从一个奇点开始。为了技巧上的原因，彭罗斯定理需要宇宙在空间上是无限的条件。于是，在实质上，我能用它来证明，只有当宇宙膨胀得快到足以避免重新坍缩时（因为只有那些弗里德曼模型才是空间无限的），才一定存在一个奇点。

在随后的几年中，我发展了新的数学技巧，从用于证明奇点一定发生的定理中除去了这个和其他技术上的条件。最后的结果是1970年彭罗斯和我的合作论文。那篇论文最后证明了，假定广义相对论是正确的，而且宇宙包含着我们观测到的这么多物质，则过去一定有过一

个大爆炸奇点。我们的工作遭遇到许多的反对，部分来自苏联人，由于他们对马克思主义科学决定论的信仰；另一部分来自某些人，他们认为整个奇点的观念是不一致的，并糟蹋了爱因斯坦理论的完美。然而，人实在不能辩赢数学定理。所以我们的工作最终被广泛接受，现在几乎每个人都假定宇宙是从一个大爆炸奇点起始的。颇具讽刺意味的是，现在我改变了想法，试图去说服其他物理学家，事实上在宇宙的开端并没有奇点——正如我们将要看到的，一旦考虑了量子效应，奇点就会消失。

我们在这一章已经看到，在不到半个世纪的时间里，人们几千年来形成的宇宙观被转变了。哈勃关于宇宙膨胀的发现，以及关于我们自己的行星在茫茫宇宙中微不足道的认识，只不过是起点而已。随着实验和理论证据的积累，人们越来越清楚地认识到，宇宙在时间上必须有个开端。直到1970年，在爱因斯坦广义相对论的基础上，彭罗斯和我才证明了它。这个证明显示，广义相对论只是一个不完全的理论，它不能告诉我们宇宙是如何开始的，因为它预言，所有包括它自己在内的物理理论都在宇宙的开端失效。然而，广义相对论宣称自己只是一个部分理论，所以奇点定理真正显示的是，在极早期宇宙中一定有过一个时刻，那时宇宙是如此之小，人们不能再不理会20世纪另一个伟大的部分理论——量子力学的小尺度效应。20世纪70年代初期，我们被迫从极其巨大范围的理论理解宇宙转变到从极其微小范围的理论理解宇宙。在我们努力将这两个部分理论结合成一个单一的量子引力论之前，下面首先描述量子力学这个理论。

第四章

# 不确定性原理

科学理论，尤其是牛顿引力论的成功，使得法国科学家拉普拉斯侯爵在19世纪初论断，宇宙是完全决定论的。拉普拉斯提出，应该存在一组科学定律，只要我们知道宇宙在某一时刻的完全的状态，我们便能预言宇宙中将会发生的任一事件。例如，假定我们知道某一个时刻的太阳和行星的位置和速度，则可用牛顿定律计算出在任何其他时刻的太阳系的状态。这种情形下的决定论是显而易见的，但拉普拉斯走得更远，他假定存在着某些类似定律，它们制约其他所有事物，包括人类的行为。

很多人强烈地抵制这种科学决定论的教义，他们感到这侵犯了上帝干涉世界的自由。但直到20世纪初，这种观念仍被认为是科学的标准假定。这种信念必须被抛弃的一个最初的征兆，是由英国科学家瑞利勋爵和詹姆斯·金斯爵士做的计算。他们指出一个热的物体——例如恒星——必须以无限大的速率辐射出能量。按照当时人们相信的定律，一个热体必须在所有的频率同等地发出电磁波（诸如射电波、可见光或X射线）。例如，一个热体在每秒1万亿次波动至2万亿次波动频率之间的波发出和在每秒2万亿次波动至3万亿次波动频率之间的波同样的能量。而既然每秒波动数是无限的，这意味着辐射出的总能量也必须是无限的。

为了避免这显然荒谬的结果，德国科学家马克斯·普朗克在1900年提出，光波、X射线和其他波不能以任意的速率辐射，而只能以某种称为量子的波包发射。此外，每个量子具有确定的能量，波的频率

越高，其能量越大。这样，在足够高的频率下，辐射单个量子所需要的能量比所能得到的还要多。因此，在高频下的辐射减少了，这样物体丧失能量的速率就变成有限的了。

量子假设可以非常成功地解释所观测到的热体的辐射发射率，但直到1926年另一位德国科学家威纳·海森伯提出著名的不确定性原理之后，人们才意识到它对决定性论的含义。为了预言一个粒子未来的位置和速度，人们必须能够准确地测量它现在的位置和速度。显而易见的办法是将光照到这粒子上。一部分光波被此粒子散射开来，由此指明它的位置。然而，人们不可能将粒子的位置确定到比光的两个波峰之间距离更小的程度，所以为了精确测量粒子的位置，必须用短波长的光。可是，由普朗克的量子假设，人们不能用任意小量的光；人们至少要用一个光量子。这量子会扰动这粒子，并以一种不能预见的方式改变粒子的速度。此外，位置测量得越准确，所需的波长就越短，单个量子的能量就越大，这样粒子的速度就被扰动得越厉害。换言之，你对粒子的位置测量得越准确，你对速度的测量就越不准确，反之亦然。海森伯指出，粒子位置的不确定性乘以粒子质量再乘以速度的不确定性不能小于一个确定量，该确定量称为普朗克常量。并且，这个极限既不依赖于测量粒子位置和速度的方法，也不依赖于粒子的种类。海森伯不确定性原理是世界的一个基本的不可回避的性质。

不确定性原理对我们的世界观有非常深远的影响。甚至到了70多年之后，许多哲学家还不能充分鉴赏它，它仍然是许多争议的主题。不确定性原理使拉普拉斯的科学理论，即一个完全决定性论的宇宙模型的梦想寿终正寝：如果人们甚至不能准确地测量宇宙现在的状态，那么就肯定不能准确地预言将来的事件！我们仍然可以想像，对于一些超自然的生物，存在一组完全地决定事件的定律，这些生物能够不干扰宇宙地观测宇宙现在的状态。然而，对于我们这些芸芸众生而言，这样的宇宙模型并没有太多的兴趣。看来，最好是采用称为奥铿

剃刀的经济原理，将理论中不能被观测到的所有特征都割除掉。20世纪20年代，在不确定性原理的基础上，海森伯、厄文·薛定谔和保罗·狄拉克运用这种手段将力学重新表述成称为量子力学的新理论。在此理论中，粒子不再分别有很好定义的而又不能被观测的位置和速度。取而代之，粒子具有位置和速度的一个结合物，量子态。

一般而言，量子力学并不对一次观测预言一个单独的确定结果。取而代之，它预言一组可能发生的不同结果，并告诉我们每个结果出现的概率。也就是说，如果我们对大量类似的系统作同样的测量，每一个系统以同样的方式起始，我们将会找到测量的结果为A出现一定的次数，为B出现另一不同的次数，等等。人们可以预言结果为A或B的出现的次数的近似值，但不能对个别测量的特定结果作出预言。因而量子力学把非预见性或随机性的不可避免因素引进了科学。尽管爱因斯坦在发展这些观念时起了很大作用，但他非常强烈地反对这些。他之所以得到诺贝尔奖就是因为他对量子理论的贡献。即使这样，他也从不接受宇宙受机缘控制的观点；他的情绪可以用他著名的断言来表达："上帝不掷骰子。"然而，其他大多数科学家愿意接受量子力学，因为它和实验符合得很完美。它的的确确成为一个极其成功的理论，并成为几乎所有现代科学技术的基础。它制约着晶体管和集成电路的行为，而这些正是电子设备诸如电视、计算机的基本元件。它还是现代化学和生物学的基础。物理科学未让量子力学适当结合进去的仅有领域是引力和宇宙的大尺度结构。

虽然光是由波组成的，普朗克的量子假设告诉我们，在某些方面，它的行为似乎显现出它是由粒子组成的——它只能以波包或量子的形式发射或吸收。同样地，海森伯的不确定性原理意味着，粒子在某些方面的行为像波一样：它们没有确定的位置，而是被"抹平"成一定的几率分布。量子力学的理论是基于一个全新的数学基础之上，不再按照粒子和波来描述实际的世界；而只不过利用这些术语，来描述对世界的观测而已。这样，在量子力学中存在着波和粒子的二重性：

为了某些目的将考虑粒子成波是有用的，而为了其他目的最好将波考虑成粒子。这导致一个很重要的结果，人们可以观察到两束波或粒子之间的所谓的干涉。那也就是，一束波的波峰可以和另一束波的波谷相重合。这两束波就相互抵消，而不像人们预料的那样，叠加在一起形成更强的波（图4.1）。一个光干涉的熟知例子是，肥皂泡上经常能看到颜色。这是因为从形成泡的很薄的水膜的两边的光反射引起的。白光由所有不同波长或颜色的光波组成，在从水膜一边反射回来的具有一定波长的波的波峰和从另一边反射的波谷相重合时，对应于此波长的颜色就不在反射光中出现，所以反射光就显得五彩缤纷。

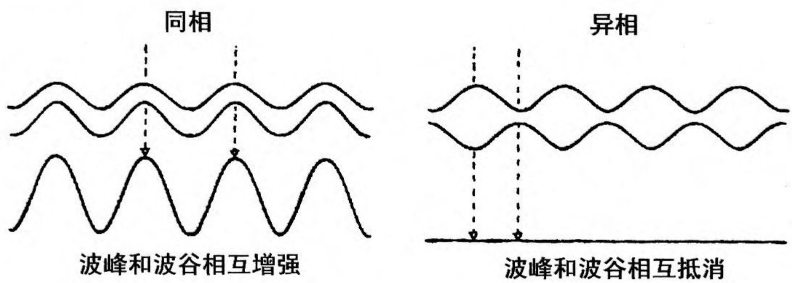



图4.1

由于量子力学引进的二重性，粒子也会产生干涉。所谓的双缝实验即是著名的例子（图4.2）。考虑一个带有两个平行狭缝的隔板，在它的一边放上一个特定颜色（即特定波长）的光源。大部分光都射在隔板上，但是一小部分光通过这两条缝。现在假定将一个屏幕放到隔板的另一边。屏幕上的任何一点都能接收到两个缝来的波。然而，一般来说，光从光源通过这两条狭缝传到屏幕上的距离是不同的。这表明，从狭缝来的光到达屏幕之时不再是相互同相的：有些地方波相互抵消，其他地方它们相互加强，结果形成有亮暗条纹的特征花样。

非常令人惊异的是，如果将光源换成粒子源，譬如具有一定速度（这表明其对应的波有确定的波长）的电子束，人们得到完全同样类

型的条纹。这显得更为古怪，因为如果只有一条裂缝，则得不到任何条纹，只不过是电子通过这屏幕的均匀分布。人们因此可能会想到，另开一条缝只不过是打到屏幕上每一点的电子数目增加而已。但是，实际上由于干涉，在某些地方反而减少了。如果在一个时刻只有一个电子被发出通过狭缝，人们会以为，每个电子只穿过这条或那条缝，这样它的行为正如只存在通过的那条缝一样——屏幕会给出一个均匀的分布。然而，实际上即便每次一个地发出电子，条纹仍然出现。因此，每个电子准是在同一时刻通过两条小缝！

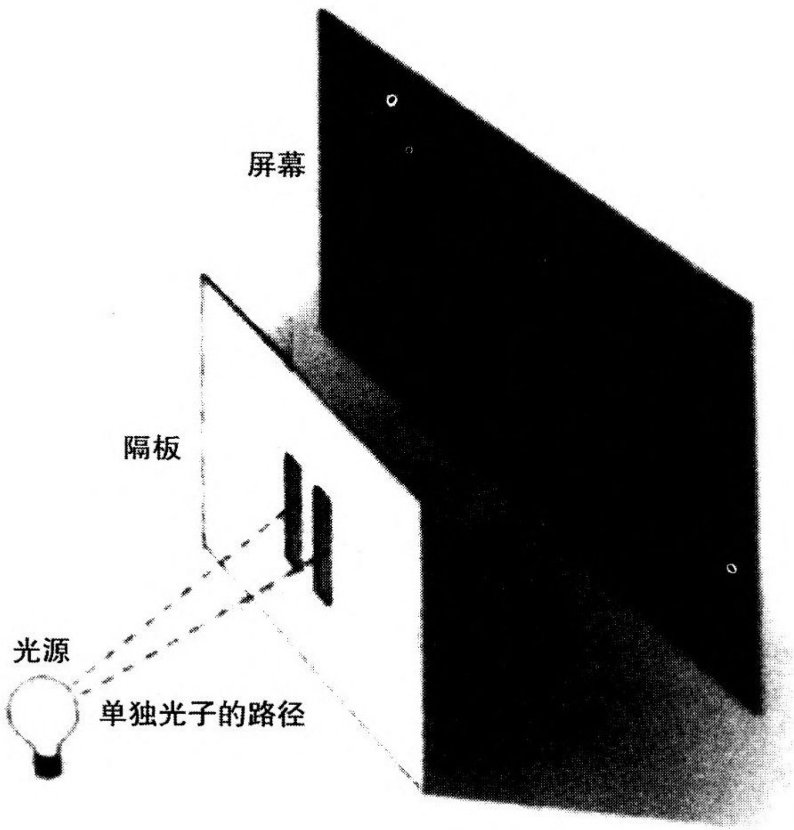



图4.2

粒子间的干涉现象，对于我们理解原子的结构至为关键，后者是

作为化学和生物的基元，以及由之组成我们和我们周围所有一切的构件。在本世纪（即20世纪——编者注）初，人们认为原子和行星围绕着太阳公转相当类似，电子（带负电荷的粒子）围绕着带正电荷的中心的核公转。人们以为正电荷和负电荷之间的吸引力维持电子的轨道，正如同行星和太阳之间的万有引力维持行星的轨道一样。麻烦在于，在量子力学之前，力学和电学的定律预言，电子会失去能量并以螺旋线的轨道落向并最终撞击到核上去。这表明原子（实际上所有的物质）都会很快地坍缩成一种非常高密度的状态。丹麦科学家尼尔斯·玻尔在1913年，为此问题找到了部分的解答。他提出，也许电子不能在离中心核任意远的地方，而只能在一些指定的距离处公转。如果我们再假定，只有一个或两个电子能在这些距离上的任一轨道上公转，因为电子除了充满最小距离和最小能量的轨道外，不能进一步向里螺旋靠近，这就解决了原子坍缩的问题。

对于最简单的原子——氢原子，这个模型给出了相当好的解释，这里只有一个电子围绕着原子核运动。但人们不清楚如何将其推广到更复杂的原子上去。并且，可允许轨道有限集合的思想似乎显得非常任意。量子力学的新理论解决了这一困难。原来一个围绕核运动的电子可被认为一个波，其波长依赖于其速度。对于一定的轨道，轨道的长度对应于整数（而不是分数）倍电子的波长。对于这些轨道，每绕一圈波峰总在同一位置，所以波就相互叠加；这些轨道对应于玻尔的可允许的轨道。然而，对于那些长度不为波长整数倍的轨道，当电子围绕着运动时，每个波峰将最终被波谷抵消；这些轨道是不允许的。

美国科学家理查德·费恩曼引入的所谓对历史求和（即路径积分）的方法是一个摹写波粒二象性的好方法。在这方法中，粒子不像在经典亦即非量子理论中那样，在时空中只有一个历史或一个路径。相反，假定粒子从A到B可走所有可能的轨道。和每个路径相关存在一对数：一个数表示波的幅度；另一个表示在周期循环中的位置（即相位）。从A走到B的几率是将所有路径的波加起来。一般说来，如果比

较一组邻近的路径，相位或周期循环中的位置会差别很大。这意味着，相应于这些轨道的波几乎都相互抵消了。然而，对于某些邻近路径的集合，它们之间的相位变化不大，这些路径的波不会抵消。这种路径对应于玻尔的允许轨道。

利用这些思想，以具体的数学形式，可以相对直截了当地计算更复杂的原子甚至分子的允许轨道。分子是由一些原子因轨道上的电子围绕不止一个原子核运动而束缚在一起形成的。由于分子的结构，以及它们之间的反应构成了化学和生物的基础，除了受不确定性原理限制之外，在原则上，量子力学允许我们预言围绕我们的几乎一切东西。（然而，实际上对一个包含稍多电子的系统需要的计算如此之复杂，以至于使我们做不到。）

看来，爱因斯坦广义相对论制约了宇宙的大尺度结构。它是所谓的经典理论；那就是说，它没有到考虑量子力学的不确定性原理，而为了和其他理论一致这是必需的。因为我们通常经历到的引力场非常弱，所以这个理论并没导致和观测的偏离。然而，早先讨论的奇点定理指出，至少在两种情形下引力场会变得非常强——黑洞和大爆炸。在这样强的场里，量子力学效应应该是非常重要的。因此，在某种意义上，经典广义相对论由于预言无限大密度的点而预示了自身的垮台，正如同经典（也就是非量子）力学由于隐含着原子必须坍缩成无限的密度，而预言自身的垮台一样。我们还没有一个完备的协调的统一广义相对论和量子力学的理论，但是我们已知这个理论所应有的一系列特征。在以下几章我们将描述这些对黑洞和大爆炸的效应。然而，此刻我们先转去介绍人类新近的尝试，他们试图将对自然界中其他力的理解合并成一个单独的统一的量子理论。

# 第五章 基本粒子和自然的力

亚里士多德相信宇宙中的所有物质由四种基本元素即土、气、火和水组成。有两种力作用在这些元素上：引力，这是指土和水往下沉的趋势；浮力，这是指气和火往上升的倾向。将宇宙的内容分割成物质和力的这种做法一直沿袭至今。

亚里士多德相信物质是连续的，也就是说，人们可以将物质无限制地分割成越来越小的小块。即人们永远不可能得到一个不可再分割下去的最小颗粒。然而几个希腊人，例如德谟克里特，则坚持物质具有固有的颗粒性，而且认为每一件东西都是由大量的各种不同类型的原子组成（原子在希腊文中的意义是“不可分的”）。争论一直持续了几个世纪，任何一方都没有任何实际的证据。但是1803年英国的化学家兼物理学家约翰·道尔顿指出，化合物总是以一定的比例结合而成的，这一事实可以用由原子聚合一起形成称作分子的个体来解释。然而，直到本世纪初这两种学派的争论才以原子论者的胜利而告终。爱因斯坦提供了其中一个重要的物理学证据。1905年，在他关于狭义相对论的著名论文发表前的几周，他在发表的另一篇文章里指出，所谓的布朗运动——浮在液体中尘埃小颗粒的无规则随机运动——可以解释为液体原子和灰尘粒子碰撞的效应。

当时就有人怀疑，这些原子终究不是不可分割的。几年前，一位剑桥大学三一学院的研究员汤姆孙演示了一种称为电子的物质粒子存在的证据。电子具有的质量比最轻原子的一千分之一还小。他使用了一种和现代电视显像管相当类似的装置：由一根红热的金属细丝发射

出电子，由于它们带负电荷，可用电场将其朝一个涂磷光物质的屏幕加速。电子一打到屏幕上就会产生一束束的闪光。人们很快即意识到，这些电子一定是从原子本身里出来的。新西兰物理学家恩斯特·卢瑟福在1911年最后证明了物质的原子确实具有内部结构：它们是由一个极其微小的带正电荷的核以及围绕着它公转的一些电子组成。他分析从放射性原子释放出的带正电荷的α粒子和原子碰撞会引起偏转的方式，从而推出这一结论。

最初，人们认为原子核是由电子和不同数量的带正电的叫做质子的粒子组成。质子是由希腊文中表达“第一”的词演化而来的，因为质子被认为是组成物质的基本单位。然而，1932年卢瑟福在剑桥的一位同事詹姆斯·查德威克发现，原子核还包含另外称为中子的粒子，中子几乎具有和质子一样大的质量但不带电荷。查德威克因这个发现获得诺贝尔奖，并被选为剑桥龚维尔和基斯学院（我即为该学院的研究员）院长。后来，他因为和其他人不和而辞去院长的职务。一群战后回来的年轻的研究员将许多已占据位置多年的老研究员选掉后，曾有过一场激烈的辩论。这是在我去以前发生的；我在这场争论尾声的1965年才加入该学院，当时另一位获诺贝尔奖的院长奈维尔·莫特爵士也因类似的争论而辞职。

直到大约30年以前，人们还以为质子和中子是“基本”粒子。但是，质子和另外的质子或电子高速碰撞的实验表明，它们事实上是由更小的粒子构成的。加州理工学院的牟雷·盖尔曼将这些粒子命名为夸克。由于对夸克的研究，他获得1969年的诺贝尔奖。此名字起源于詹姆斯·乔伊斯神秘的引语：“Three quarks for Muster Mark！”夸克这个字应发夸脱的音，但是最后的字母是k而不是t，通常和拉克（云雀）相押韵。

存在有几种不同类型的夸克——有六种“味”，这些味我们分别称之为上、下、奇、粲、底和顶。20世纪60年代起人们就知道前三种夸克，1974年才发现粲夸克，1977年和1995年分别发现底和顶夸克。

每种味都带有三种“色”，即红、绿和蓝。（必须强调，这些术语仅仅是标签：夸克比可见光的波长小得多，所以在通常意义下没有任何颜色。这只不过是现代物理学家似乎更富有想像力地命名新粒子和新现象的方式而已——他们不再让自己受限制于希腊文！）一个质子或中子由三个夸克组成，每个夸克各有一种颜色。一个质子包含两个上夸克和一个下夸克；一个中子包含两个下夸克和一个上夸克。我们可以创生由其他种类的夸克（奇、粲、底和顶）构成的粒子，但所有这些都具有大得多的质量，并非常快地衰变成质子和中子。

现在我们知道，不管是原子还是其中的质子和中子都不是不可分的。问题在于什么是真正的基本粒子——构成世界万物的最基本的构件？由于光波波长比原子的尺度大得多，我们不能期望以通常的方法去“看”一个原子的部分。我们必须用某些波长短得多的东西。正如我们在上一章所看到的，量子力学告诉我们，实际上所有粒子都是波，粒子的能量越高，则其对应的波的波长越短。所以，我们能对这个问题给出的最好的回答，取决于我们装置中的粒子能量有多高，因为这决定了我们能看到的尺度有多小。这些粒子的能量通常用叫做电子伏特的单位来测量。（在汤姆孙的电子实验中，我们看到他用一个电场去加速电子，一个电子从一个伏特的电场所得到的能量即是一个电子伏特。）19世纪，当人们知道如何去使用的粒子能量只是由化学反应——诸如燃烧——产生的几个电子伏特的低能量时，大家以为原子即是最小的单位。在卢瑟福的实验中，α粒子具有几百万电子伏特的能量。更晚的时代，我们获悉如何使用电磁场给粒子提供首先是几百万，然后是几十亿电子伏特的能量。这样我们知道，30年之前以为是“基本”的粒子，事实上是由更小的粒子组成。如果我们利用更高的能量时，是否会发现这些粒子是由更小的粒子组成的呢？这一定是可能的。但我们确实有一些理论上的原因，相信我们已经拥有，或者说接近拥有自然的终极构件的知识。

用上一章讨论的波粒二象性，包括光和引力的宇宙中的一切都能

以粒子来描述。这些粒子有一种称为自旋的性质。考虑自旋的一个方法是将粒子想象成围绕着一个轴自转的小陀螺。然而，这可能会引起误会，因为量子力学告诉我们，粒子并没有任何轮廓分明的轴。粒子的自旋真正告诉我们的是，从不同的方向看粒子是什么样子的。一个自旋为0的粒子像一个点：从任何方向看都一样〔图5.1（i）〕。另一方面，自旋为1的粒子像一个箭头：从不同方向看是不同的〔图5.1（ii）〕。只有把它转过一整圈（360°）时，这粒子才显得一样。自旋为2的粒子像个双头的箭头〔图5.1（iii）〕：只要把它转过半圈（180°），它看起来便一样。类似地，把更高自旋的粒子转了整圈的更小的部分后，它看起来便一样。所有这一切都是这样的直截了当，但惊人的事实是，把有些粒子转过一圈后，它仍然显得不同：你必须使其转两整圈！这样的粒子就说具有1/2的自旋。

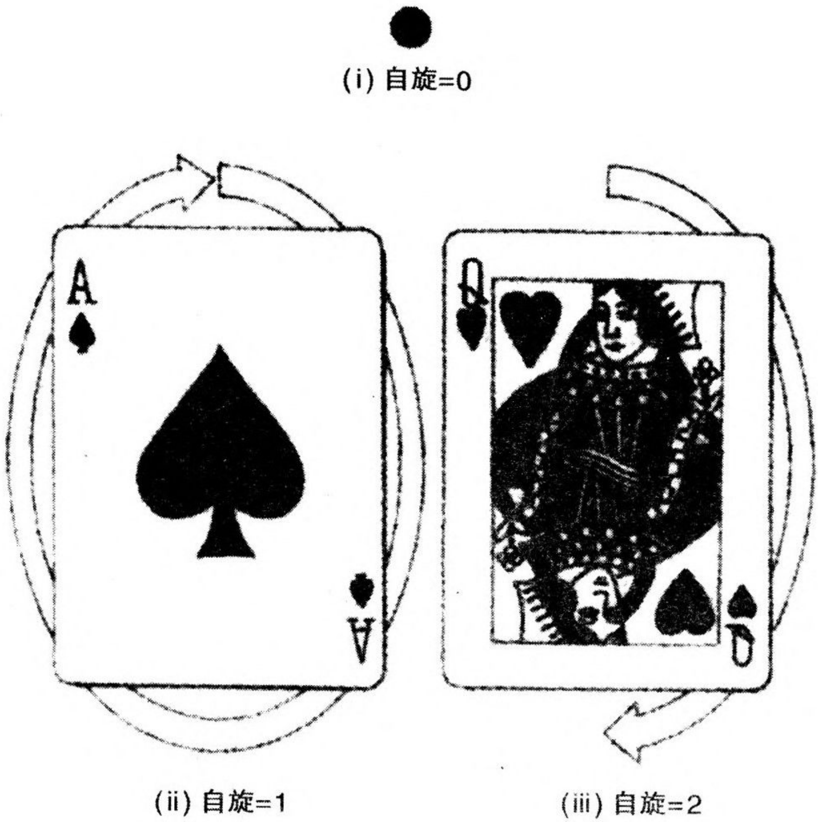



图5.1

宇宙间所有已知的粒子可以分成两组：自旋为1/2的粒子，它们组成宇宙中的物质；自旋为0、1和2的粒子，正如我们将要看到的，它们在物质粒子之间产生力。物质粒子服从所谓的泡利不相容原理。这是奥地利物理学家沃尔夫冈·泡利在1925年发现的，他因此而获得1945年的诺贝尔奖。他是个原型的理论物理学家，有人这样说，他的存在甚至会使同一城市里的实验出毛病！泡利不相容原理是说，两个类似的粒子不能存在于相同的态中，也就是说，在不确定性原理给出的限制下，它们不能同时具有相同的位置和速度。不相容原理是非常关键的，因为它解释了为何物质粒子，在自旋为O、1和2的粒子产生

的力的影响下，不会坍缩成密度非常高的状态的原因：如果物质粒子几乎处在相同的位置，则它们必须有不同的速度，这意味着它们不会长时间存在于相同的位置。如果世界在没有不相容原理的情形下创生，夸克将不会形成分离的轮廓分明的质子和中子，进而这些也不可能和电子形成分离的轮廓分明的原子。它们全部都会坍缩形成大致均匀的稠密的“汤”。

直到保罗·狄拉克在1928年提出一个理论，人们才对电子和其他自旋1/2的粒子有了正确的理解。狄拉克后来被选为剑桥的卢卡斯数学教授（牛顿曾经担任这一教席，目前我担任这一职务）。狄拉克理论是第一种既和量子力学又和狭义相对论相一致的理论。它在数学上解释了为何电子具有1/2的自旋，也即为什么将其转一整圈不能、而转两整圈才能使它显得一样。它还预言了电子必须有它的配偶——反电子或正电子。1932年正电子的发现证实了狄拉克的理论，他因此获得了1933年的诺贝尔奖。现在我们知道，任何粒子都有会和它相湮灭的反粒子。（对于携带力的粒子，反粒子即为其自身）。也可能存在由反粒子构成的整个反世界和反人。然而，如果你遇到了反自身，注意不要握手！否则，你们两人都会在一个巨大的闪光中消失殆尽。为何我们周围的粒子比反粒子多得多是一个极端重要的问题，我将会在本章的后部分回到这问题上来。

在量子力学中，所有物质粒子之间的力或相互作用都认为是由自旋为整数0、1或2的粒子携带。所发生的是，物质粒子——譬如电子或夸克——发出携带力的粒子。这个发射引起的反弹，改变了物质粒子的速度。携带力的粒子然后和另一个物质粒子碰撞并且被吸收。这碰撞改变了第二个粒子的速度，正如同这两个物质粒子之间存在过一个力。携带力的粒子不服从泡利不相容原理，这是它们的一个重要的性质。这表明它们能被交换的数目不受限制，这样它们就可以引起很强的力。然而，如果携带力的粒子具有很大的质量，则在大距离上产生和交换它们就会很困难。这样，它们所携带的力只能是短程的。另

一方面，如果携带力的粒子本身质量为零，力就是长程的了。因为在物质粒子之间交换的携带力的粒子，不像“实”粒子那样可以用粒子探测器检测到，所以称为虚粒子。然而，因为它们具有可测量的效应，即它们引起了物质粒子之间的力，所以我们知道它们存在。自旋为0、1或2的粒子在某些情况下也作为实粒子存在，这时它们可以被直接探测到。对我们而言，此刻它们就呈现出经典物理学家称为波动形式，例如光波和引力波的东西。当物质粒子以交换携带力的虚粒子的形式而相互作用时，它们有时就可以被发射出来。（例如，两个电子之间的电排斥力是由于交换虚光子所致，这些虚光子永远不可能被检测出来；但是如果一个电子从另一个电子边穿过，则可以放出实光子，它作为光波而被我们探测到。）

携带力的粒子按照其强度以及与其相互作用的粒子可以分成四个种类。必须强调指出，这种将力划分成四种是人为的；它仅仅是为了便于建立部分理论，而并不别具深意。大部分物理学家希望最终找到一个统一理论，该理论将四种力解释为一个单独的力的不同方面。确实，许多人认为这是当代物理学的首要目标。最近，将四种力中的三种统一起来已经有了成功的端倪——我将在这一章描述这些内容。而关于统一余下的另一种力即引力的问题将留到以后。

第一种力是引力，这种力是万有的，也就是说，每一个粒子都因它的质量或能量而感受到引力。引力比其他三种力都弱得多。它是如此之弱，它若不具有两个特别的性质，我们根本就不可能注意到：它能作用到大距离去，以及它总是吸引的。这意味着，在像地球和太阳这样两个巨大的物体中，单独粒子之间的非常弱的引力能都叠加起来而产生相当大的力量。其他三种力要么是短程的，要么时而吸引时而排斥，所以它们倾向于相互抵消。以量子力学的方法来看待引力场，人们把两个物质粒子之间的力描述成由称作引力子的自旋为2的粒子携带的。它自身没有质量，所以携带的力是长程的。太阳和地球之间的引力可以归结为构成这两个物体的粒子之间的引力子交换。虽然所

交换的粒子是虚的，它们确实产生了可测量的效应——它们使地球围绕着太阳公转！实引力子构成了经典物理学家称之为引力波的东西，它是如此之弱——并且要探测到它是如此之困难，以至于还从来未被观测到过。

另一种力是电磁力。它作用于带电荷的粒子（例如电子和夸克）之间，但不和不带电荷的粒子（例如引力子）相互作用。它比引力强得多：两个电子之间的电磁力比引力大约大10　0亿亿亿亿亿（在1后面有4　2个O）倍。然而，存在两种电荷——正电荷和负电荷。同种电荷之间的力是相互排斥的，而异种电荷之间的力则是相互吸引的。一个大的物体，譬如地球或太阳，包含了几乎等量的正电荷和负电荷。这样，由于单独粒子之间的吸引力和排斥力几乎全被抵消了，因此两个物体之间净的电磁力非常小。然而，电磁力在原子和分子的小尺度下起主要作用。在带负电的电子和带正电的核中的质子之间的电磁力使得电子围绕着原子的核公转，正如同引力使得地球围绕着太阳公转一样。人们将电磁吸引力描绘成是由于交换大量称作光子的无质量的自旋为1的虚粒子引起的。重复一下，这里交换的光子是虚粒子。但是，电子从一个允许轨道转变到另一个离核更近的允许轨道时，释放能量并且发射出实光子——如果其波长适当，则作为可见光可被肉眼观察到，或可用诸如照相底版的光子探测器观察到。同样，如果一个光子和原子相碰撞，可将电子从离核较近的允许轨道移动到较远的轨道。这样光子的能量被消耗掉，它也就被吸收了。

第三种力称为弱核力。它负责放射性现象，并只作用于自旋为1/2的所有物质粒子，而对诸如光子、引力子等自旋为0、1或2的粒子不起作用。直到1967年伦敦帝国学院的阿伯达斯·萨拉姆和哈佛的史蒂芬·温伯格提出了弱作用和电磁作用的统一理论后，弱作用才被很好地理解。此举在物理学界所引起的震动，可与大约100年前麦克斯韦统一电学和磁学相提并论。他们提出，除了光子，还存在其他3个自旋为1的被统称作重矢量玻色子的粒子，它们携带弱力。它们称作W

＋（W正）、W-（W负）和Z（Z零），每一个都具有大约100吉电子伏的质量（1吉电子伏为10亿电子伏）。温伯格-萨拉姆理论展现了称作对称自发破缺的性质。这意味着，在低能量下一些看起来完全不同的粒子，事实上发现都只是同一种粒子处于不同的状态。所有这些粒子在高能量下都有相似的行为。这个效应和轮赌盘上的轮赌球的行为相类似。在高能量下（当这轮子转得很快时），这球的行为基本上只有一个方式——即不断地滚动着。但是随着轮子变慢下来，球的能量减小，最终球就陷到轮子上的37个槽中的一个里去。换言之，在低能下球可以存在于37种不同的状态。如果由于某种原因，我们只能在低能下观察球，我们就会以为存在37种不同类型的球！

在温伯格·萨拉姆理论中，当能量远远超过100吉电子伏时，这3种新粒子和光子都以相似的方式行为。但是，大部分正常情况下粒子能量要比这低，粒子之间的对称被破坏了。W＋、W-和Z。得到了大的质量，使之携带的力变成非常短程。萨拉姆和温伯格提出此理论时，很少人相信他们，因为加速器还未强大到将粒子加速到产生实的W＋、W-和Z粒子所需的100吉电子伏的能量。但在此后的十几年里，在较低能量下这个理论的其他预言和实验符合得这样好，使他们和也在哈佛的谢尔登·格拉肖一起获得1979年的诺贝尔物理学奖。格拉肖提出过一个类似的统一电磁和弱作用的理论。由于1983年在CERN（欧洲核子研究中心）发现了具有被正确预言的质量和其他性质的光子的3个有质量的伴侣，使得诺贝尔委员会避免了犯错误的难堪。领导几百名物理学家作出此发现的卡罗·鲁比亚和开发了被使用的反物质储藏系统的CERN工程师西蒙·范德·米尔分享了1984年的诺贝尔奖。（除非你已经是巅峰人物，当今要在实验物理学上留下痕迹极其困难！）

第四种力是强核力。它将质子和中子中的夸克束缚在一起，并将原子核中的质子和中子束缚在一起。人们相信，称为胶子的另一种自旋为1的粒子携带强作用力。它只能与自身以及与夸克相互作用。强

核力具有一种称为禁闭的古怪性质：它总是把粒子束缚成不带颜色的结合体。由于夸克有颜色（红、绿或蓝），人们不能得到单独的夸克自身。相反，一个红夸克必须用一串胶子和一个绿夸克以及一个蓝夸克连接在一起（红＋绿＋蓝＝白）。这样的三胞胎构成了一个质子或中子。其他的可能性是由一个夸克和一个反夸克组成的对（红＋反红，或绿＋反绿，或蓝＋反蓝＝白）。这样的结合体构成了称为介子的粒子。介子是不稳定的，因为夸克和反夸克会相互湮灭，而产生电子和其他粒子。类似地，由于胶子也有颜色，色禁闭使得人们不可能得到单独的胶子自身。相反，人们所能得到的胶子的团，其叠加起来的颜色必须是白的。这样的团形成了称为胶球的不稳定粒子。

色禁闭使得人们观察不到一个孤立的夸克或胶子，这事实使得将夸克和胶子当作粒子的整个见解看起来有点玄学的味道。然而，强核力还有一种叫做渐近自由的性质，它使得夸克和胶子成为意义明确的概念。在正常能量下，强核力确实很强，它将夸克紧紧地捆在一起。但是，大型粒子加速器的实验指出，强作用力在高能量下变得弱得多，夸克和胶子的行为就几乎像自由粒子那样。图5.2是一张显示一个高能质子和一个高能反质子碰撞的照片。

统一电磁力和弱核力的成功，使人们多次试图将这两种力和强核力合并在所谓的大统一理论（或GUT）之中。这名字相当夸张：得到的理论并不那么辉煌，也没能将全部力都统一进去，因为它并不包含引力。它们也不是真正完整的理论，因为它们包含了许多不能从这理论中预言而必须人为选择去适合实验的参数。尽管如此，它们可能是朝着完备的统一理论推进的一步。GUT的基本思想是这样：正如前面提到的，在高能量下强核力变弱了；另一方面，不是渐近自由的电磁力和弱力在高能量下变强了。在某个非常高的叫做大统一能量的能量下，这3种力都具有同样的强度，并因此可看成一个单独的力的不同方面。在这能量下，GUT还预言了自旋为1/2的不同物质粒子（如夸克和电子）也会根本上都变成一样，这样导致了另一种统一。



图5.2　一个质子和一个反质子在高能下碰撞，产生了一对几乎自由的夸克

大统一能量的数值还知道得不太清楚，可能至少有1000万亿吉电子伏特。而目前粒子加速器只能使大致能量为100吉电子伏的粒子相碰撞，而计划建造的机器的能量可升到几千吉电子伏。要建造足以将粒子加速到大统一能量的机器，其体积必须和太阳系一样大——这在现代经济环境下不太可能做到。因此，不可能在实验室里直接检验大统一理论。然而，如同在弱电统一理论中那样，我们可以检验它在低能量下的推论。

其中最有趣的预言是，构成通常物质的大部分质量的质子能够自发衰变成诸如反电子之类更轻的粒子。之所以可能，其原因在于，在大统一能量下，夸克和反电子之间没有本质的不同。在正常情况下一个质子中的三个夸克没有足够能量转变成反电子，由于不确定性原理意味着质子中夸克的能量不可能严格不变，其中一个夸克会非常偶然地获得足够能量进行这种转变。这样质子就要衰变。夸克要得到足够能量的概率是如此之低，至少要等待100万亿亿亿年（1后面跟30个0）才能有1次。这比宇宙从大爆炸以来的年龄（大约100亿年——1

后面跟10个0）要长得多了。因此，人们会认为不可能在实验上检测到质子自发衰变的可能性。然而，人们可以观察包含极大数量质子的大量物质，以增加检测衰变的机会。（譬如，如果观察的对象含有1后面跟31个0个质子，按照最简单的GUT，可以预料在1年内应能看到多于一次的质子衰变）。

人们进行了一系列实验，可惜没有得到任何质子或中子衰变的确实证据。有一个实验是在俄亥俄的莫尔顿盐矿里进行的（为了避免其他因宇宙射线引起的会和质子衰变相混淆的事件发生），用了8000吨水。由于在实验中没有观测到自发的质子衰变，因此可以估算出，可能的质子寿命至少应为1000万亿亿亿（1后面跟31个0）年。这比简单的大统一理论所预言的寿命更长。然而，一些更精致的大统一理论预言的寿命比这更长，因此需要用更灵敏的手段对甚至更大量的物质进行检验。

尽管观测质子的自发衰变非常困难，但很可能正由于这相反的过程，即质子，或更简单地说，夸克的产生导致了我们的存在。它们是从宇宙开初的可以想像的最自然的方式——一夸克并不比反夸克更多的状态下产生的。地球上的物质主要是由质子和中子，进而由夸克构成。除了少数由物理学家在大型粒子加速器中产生的以外，不存在由反夸克构成的反质子和反中子。我们从宇宙线中得到的证据表明，我们星系中的所有物质也是这样：除了少数当粒子和反粒子对进行高能碰撞时产生的以外，没有发现反质子和反中子。如果在我们星系中有很大区域的反物质，则可以预料，在正反物质的边界会观测到大量的辐射。许多粒子在那里和它们的反粒子相碰撞、相互湮灭并释放出高能辐射。

我们没有直接的证据，表明其他星系中的物质是由质子、中子还是由反质子、反中子构成，但两者必居其一，在单一的宇宙中不能有混合，否则，我们又会观察到大量由湮灭产生的辐射。因此，我们相信，所有的星系是由夸克而不是反夸克构成；看来，一些星系为物

质，而另一些星系为反物质也是难以置信的。

为什么夸克比反夸克多这么多？为何它们的数目不相等？这数目有所不同肯定使我们交了好运，否则，早期宇宙中它们势必已经相互湮灭了，只余下一个充满辐射而几乎没有物质的宇宙。因此，后来也就不会有人类生命赖以发展的星系、恒星和行星。庆幸的是，大统一理论可以解释，尽管甚至刚开始时两者数量相等，为何现在宇宙中夸克比反夸克多。正如我们已经看到的，大统一理论允许夸克变成高能下的反电子。它们也允许相反的过程，反夸克变成电子，电子和反电子变成反夸克和夸克。在极早期宇宙有一时期是如此之热，粒子能量高到足以发生这些转变。但是，它为何使夸克比反夸克多呢？原因在于，物理定律对于粒子和反粒子不是完全相同的。

直到1956年人们都相信，物理定律分别服从三个叫做C、P和T的对称。C（电荷）对称的意义是，定律对于粒子和反粒子是相同的；P（宇称）对称的意义是，定律对于任何情景和它的镜像（右手方向自旋的粒子的镜像变成了左手方向自旋的粒子）是相同的；T（时间）对称的意义是，如果你颠倒所有粒子和反粒子的运动方向，系统应回到早先的那样；换言之，定律对于前进或后退的时间方向是一样的。1956年，两位美国物理学家李政道和杨振宁提出弱作用实际上不服从P对称。换言之，弱力使得宇宙和宇宙的镜像以不同的方式发展。同一年，他们的一位同事吴健雄证明了他们的预言是正确的。她把放射性原子的核排列在磁场中，使它们的自旋方向一致。实验表明，在一个方向比另一方向发射出得更多电子。次年，李和杨为此获得诺贝尔奖。人们还发现弱作用不服从C对称，即是说，它使得由反粒子构成的宇宙以和我们的宇宙不同的方式行为。尽管如此，弱力似乎确实服从CP联合对称。也就是说，如果每个粒子都用其反粒子来取代，则由此构成的宇宙的镜像和原来的宇宙以同样的方式发展！然而，1964年，还是两个美国人——J·W·克罗宁和瓦尔·费兹——发现，在某种称为K介子的衰变中，甚至连CP对称也不服从。1980年，克罗

宁和费兹最终由于他们的研究而获得诺贝尔奖。（很多奖是因为显示宇宙不像我们曾经想像的那么简单而授予的！）

有一个数学定理说，任何服从量子力学和相对论的理论必须服从CPT联合对称。换言之，如果同时用反粒子来置换粒子，取镜像还有时间反演，则宇宙的行为必须是一样的。但是，克罗宁和费兹指出，如果仅仅用反粒子来取代粒子，并且采用镜像，但不反演时间方向，则宇宙的行为不相同。所以，如果人们反演时间方向，物理学定律必须改变——它们不服从T对称。

早期宇宙肯定是不服从T对称的：随着时间前进，宇宙膨胀——如果它往后倒退，则宇宙收缩。而且，由于存在着不服从T对称的力，因此当宇宙膨胀时，相对于将电子变成反夸克，这些力将更多的反电子变成夸克。然后，随着宇宙膨胀并冷却下来，反夸克就和夸克湮灭，但由于已有的夸克比反夸克多，少量过剩的夸克就留下来。正是它们构成我们今天看到的物质，由这些物质构成了我们自身。这样，我们自身之存在可认为是大统一理论的证实，哪怕仅仅是定性的而已；但此预言的不确定性到了这种程度，以至于我们不能知道在湮灭之后余下的夸克数目，甚至不知是夸克还是反夸克余下。（然而，如果是反夸克多余留下，我们可以简单地把反夸克称为夸克，夸克称为反夸克。）

大统一理论不包括引力。在我们处理基本粒子或原子问题时这关系不大，因为引力是如此之弱，通常可以忽略它的效应。然而，它的作用既是长程的，又总是吸引的事实，表明它的所有效应是叠加的。所以，对于足够大量的物质粒子，引力会比其他所有的力都更重要。这就是为什么正是引力决定了宇宙的演化的缘故。甚至对于恒星大小的物体，引力的吸引会超过所有其他的力，并使恒星坍缩。我在70年代的工作集中于研究黑洞。黑洞就是由这种恒星的坍缩和围绕它们的强大的引力场产生的。正是黑洞研究给出了量子力学和广义相对论如何相互影响的第一个暗示——亦即尚未成功的量子引力论形态的一

瞥。

# 第六章 黑洞

黑洞这一术语是不久以前才出现的。1969年美国科学家约翰·惠勒，为了形象地描述至少可回溯到200年前的一个观念时，杜撰了这个名词。那时候，共有两种光理论：一种是牛顿赞成的光的微粒说；另一种是光由波构成的波动说。我们现在知道，这两者在实际上都是正确的。由于量子力学的波粒二象性，光既可认为是波，也可认为是粒子。在光的波动说中，不清楚光对引力如何响应。但是如果光是由粒子组成的，人们可以预料，它们正如同炮弹、火箭和行星一样受引力的影响。人们起先以为，光粒子无限快地运动，所以引力不可能使之缓慢下来，但是罗默关于光以有限速度行进的发现意味着，引力对之可有重要效应。

1783年，剑桥的学监约翰·米歇尔在这个假定的基础上，于《伦敦皇家学会哲学学报》上发表了一篇文章。他指出，一个质量足够大并足够致密的恒星会有如此强大的引力场，甚至连光线都不能逃逸：任何从恒星表面发出的光，在还没到达远处前就会被恒星的引力吸引回来。米歇尔暗示，可能存在大量这样的恒星，虽然由于从它们那里发出的光不会到达我们这里，我们不能看到它们；但是我们仍然可以感到它们引力的吸引。这正是我们现在称为黑洞的物体。它是名副其实的——在空间中的黑的空洞。几年之后，法国科学家拉普拉斯侯爵显然独自地提出了和米歇尔类似的观念。非常有趣的是，拉普拉斯只将此观点纳入他的《世界系统》一书的第一版和第二版中，而在以后的版本中将其删去；也许他认为这是一个愚蠢的观念。（还有，光的微粒说在19世纪变得不时髦了；似乎一切都可以以波动理论来解释，

而按照波动理论，不清楚光究竟是否受到引力的影响。）

事实上，因为光速是固定的，所以在牛顿引力论中将光类似炮弹那样处理不很协调。（从地面发射上天的炮弹被引力减速，最后停止上升并折回地面；然而，一个光子必须以不变的速度继续向上，那么，牛顿引力如何影响光呢？）直到1915年爱因斯坦提出广义相对论，才得到引力如何影响光的协调理论。甚至又过了很长时间，人们才理解这个理论对大质量恒星的含意。

为了理解黑洞是如何形成的，我们首先需要理解恒星的生命周期。起初，大量的气体（绝大部分为氢）受自身的引力吸引，而开始向自身坍缩而形成恒星。当它收缩时，气体原子越来越频繁地以越来越大的速度相互碰撞——气体的温度上升。最后，气体变得如此之热，以至于当氢原子碰撞时，它们不再弹开而是聚合形成氦。如同一个受控氢弹爆炸，反应中释放出来的热使得恒星发光。这附加的热又使气体的压力升高，直到它足以平衡引力的吸引，这时气体停止收缩。这有一点像气球——内部气压试图使气球膨胀，橡皮的张力试图使气球收缩，它们之间存在一个平衡。从核反应发出的热和引力吸引的平衡，使恒星在很长时间内维持这种平衡。然而，恒星最终会耗尽它的氢和其他核燃料。貌似大谬，其实不然的是，恒星初始的燃料越多，它则被越快燃尽。这是因为恒星的质量越大，它就必须越热才足以抵抗引力。而它越热，它的燃料就被耗得越快。我们的太阳大概足够再燃烧50多亿年，但是质量更大的恒星可以在1亿年这么短的时间内耗尽其燃料，这个时间尺度比宇宙的年龄短得多了。当恒星耗尽了燃料，它开始变冷并收缩。随后发生的情况只有等到20世纪20年代末才首次被人们理解。

1928年，一位印度研究生——萨拉玛尼安·昌德拉塞卡——乘船来英国剑桥跟英国天文学家兼广义相对论家阿瑟·爱丁顿爵士学习。（据记载，在20世纪20年代初，有一位记者告诉爱丁顿，说他听说世界上只有三个人能理解广义相对论。爱丁顿停顿了一下，然后回

答："我正在想这第三个人是谁？"）在从印度来英国的旅途中，昌德拉塞卡算出了在耗尽所有燃料之后，多大的恒星仍然可以对抗自己的引力而维持本身。这个思想是说：当恒星变小时，物质粒子相互靠得非常近，而按照泡利不相容原理，它们必须有非常不同的速度。这使得它们相互散开并企图使恒星膨胀。因此，一颗恒星可因引力的吸引和不相容原理引起的排斥达到的平衡，而保持其半径不变，正如同在它的生命的早期引力被热平衡一样。

然而，昌德拉塞卡意识到，不相容原理所能提供的排斥力有一个极限。相对论把恒星中的粒子的最大速度差限制为光速。这意味着，当恒星变得足够密集之时，由不相容原理引起的排斥力就会比引力的作用小。昌德拉塞卡计算出，一个质量比大约太阳质量一倍半还大的冷的恒星不能维持本身以抵抗自己的引力。（这质量现在称为昌德拉塞卡极限。）苏联科学家列夫·达维多维奇·朗道差不多同时得到了类似的发现。

这对大质量恒星的最终归宿具有重大的意义。如果一颗恒星的质量比昌德拉塞卡极限小，它最后会停止收缩，并且变成一种可能的终态即"白矮星"。白矮星的半径为几千英里，密度为每立方英寸几百吨。白矮星是由它物质中电子之间的不相容原理排斥力支持的。我们观察到大量这样的白矮星。围绕着天狼星转动的那颗是最早被发现的白矮星中的一个，天狼星是夜空中最亮的恒星。

朗道指出，恒星还存在另一种可能的终态。其极限质量大约也为太阳质量的一倍或二倍，但是其体积甚至比白矮星还小得多。这些恒星是由中子和质子之间，而不是电子之间的不相容原理排斥力支持的。所以它们叫做中子星。它们的半径只有10英里左右，密度为每立方英寸几亿吨。在第一次预言中子星时，没有任何方法去观察它。实际上，它们很久以后才被探测到。

另一方面，质量比昌德拉塞卡极限还大的恒星在耗尽其燃料时，

会出现一个很大的问题。在某种情形下，它们会爆炸或设法抛出足够的物质，使它们的质量减小到极限之下，以避免灾难性的引力坍缩。但是很难令人相信，不管恒星有多大，这总会发生。怎么知道它一定损失重量呢？即使每个恒星都设法失去足够多的质量以避免坍缩，如果你把更多的质量加在白矮星或中子星上，以使之超过极限，将会发生什么？它会坍缩到无限密度吗？爱丁顿为此感到震惊，他拒绝相信昌德拉塞卡的结果。爱丁顿认为，一颗恒星是根本不可能坍缩成一点的。这是大多数科学家的观点：爱因斯坦自己写了一篇论文，宣布恒星的体积不会收缩为零。其他科学家，尤其是他以前的老师，恒星结构的主要权威——爱丁顿的敌意使昌德拉塞卡放弃了这方面的工作，而转去研究诸如恒星团运动等其他天文学问题。然而，他之所以获得1983年诺贝尔奖，至少部分原因在于他早年所做的关于冷恒星的质量极限的工作。

昌德拉塞卡指出，不相容原理不能够阻止质量大于昌德拉塞卡极限的恒星发生坍缩。但是，根据广义相对论，这样的恒星会发生什么情况呢？1939年一位美国的年轻人罗伯特·奥本海默首次解决了这个问题。然而，他所获得的结果表明，用当时的望远镜去检测不会有任何观测结果。以后，第二次世界大战插入，奥本海默本人非常专心地从事原子弹研制。战后，由于大部分科学家被吸引到原子和原子核尺度的物理中去，因而大部分人忘记了引力坍缩的问题。但在20世纪60年代，现代技术的应用使得天文观测范围和数量大大增加，这重新激起人们对天文学和宇宙学的大尺度问题的兴趣。奥本海默的工作被一些人重新发现并推广。

现在，我们从奥本海默的工作中得到一幅这样的图象：恒星的引力场改变了光线在时空中的路径，使之和如果没有恒星情况下的路径不一样。光锥是表示闪光从其顶端发出后在时空中传播的路径。光锥在恒星表面附近稍微向内弯折。在日食时观察从遥远恒星发出的光线，可以看到这种偏折现象。随着恒星收缩，其表面的引力场变得更

强大，而光锥向内偏折得更多。这使得光线从恒星逃逸变得更为困难，对于远处的观察者而言，光线变得更黯淡更红。最后，当恒星收缩到某一临界半径时，表面上的引力场变得如此之强，使得光锥向内偏折得这么厉害，以至于光线再也逃逸不出去（图6.1）。根据相对论，没有东西能行进得比光还快。这样，如果光都逃逸不出来，其他东西更不可能：所有东西都会被引力场拉回去。这样，存在一个事件的集合或时空区域，光或任何东西都不可能从该区域逃逸而到达远处的观察者。现在我们将这区域称作黑洞，将其边界称作事件视界，而它和刚好不能从黑洞逃逸的光线的那些路径相重合。

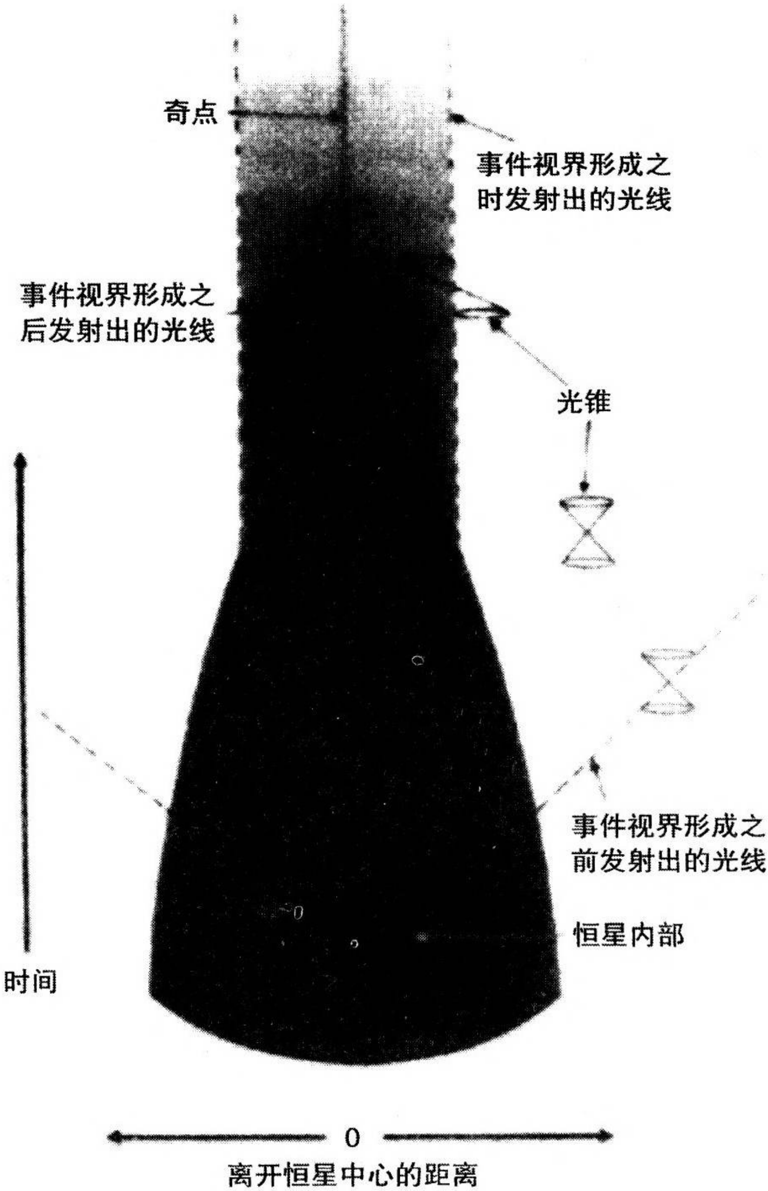



图6.1

如果你观察一个恒星坍缩并形成黑洞时，为了理解你所看到的情况，切记在相对论中没有绝对时间。每个观测者都有自己的时间测

量。由于恒星的引力场，在恒星上某人的时间将和在远处某人的时间不同。假定在坍缩星表面有一无畏的航天员和恒星一起向内坍缩。他按照自己的表，每一秒钟发一信号到一个围绕着该恒星转动的航天飞船上去。在他的表的某一时刻，譬如11点钟，恒星刚好收缩到它的临界半径以下，此时引力场强大到没有任何东西可以逃逸出去，他的信号再也不能传到航天飞船了。随着11点趋近，他的伙伴从航天飞船上观看会发现，从该航天员发来的一串信号的时间间隔越变越长。但是这个效应在10点59分59秒之前是非常微小的。在收到10点59分58秒和10点59分59秒发出的两个信号之间，他们只需等待比1秒钟稍长一点的时间，然而他们必须为11点发出的信号等待无限长的时间。按照航天员的手表，光波是在10点59分59秒和11点之间由恒星表面发出；从航天飞船上看，那光波被散开到无限长的时间间隔里。在航天飞船上这一串光波来临的时间间隔变得越来越长，所以从恒星来的光显得越来越红、越来越淡，最后，该恒星变得如此之朦胧，以至于从航天飞船上再也看不见它：所余下的一切只是空间中的一个黑洞。不过，此恒星继续以同样的引力作用到航天飞船上，使飞船继续围绕着形成的黑洞旋转。但是由于以下的问题，上述场景不是完全现实的。一个人离开恒星越远则引力越弱，所以作用在这位无畏的航天员脚上的引力总比作用到他头上的大。在恒星还未收缩到临界半径而形成事件视界之前，这力的差别就足以将我们的航天员拉成意大利面条那样，甚至将他撕裂！然而我们相信，在宇宙中存在大得多的天体，譬如星系的中心区域，它们遭受到引力坍缩而产生黑洞；一位在这样的物体上面的航天员在黑洞形成之前不会被撕开。事实上，当他到达临界半径时，不会有任何异样的感觉，甚至在通过永不回返的那一点时，都没注意到它。然而，随着这区域继续坍缩，只要在几个钟头之内，作用到他头上和脚上的引力之差会变得如此之大，以至于再将其撕裂。

罗杰·彭罗斯和我在1965年和1970年之间的研究指出，根据广义相对论，在黑洞中必然存在密度和时空曲率无限大的奇点。这和时间

开端时的大爆炸相当类似，只不过它是一个坍缩物体和航天员的时间终点而已。在此奇点，科学定律和我们预言将来的能力都崩溃了。然而，任何留在黑洞之外的观察者，将不会受到可预见性失效的影响，因为从奇点出发的，不管是光还是任何其他信号，都不能到达他那儿。这个非凡的事实导致罗杰·彭罗斯提出了宇宙监督假想，它可以被意译为："上帝憎恶裸奇点。"换言之，由引力坍缩所产生的奇点只能发生在像黑洞这样的地方，它在那里被事件视界体面地遮住而不被外界看见。严格地讲，这就是所谓弱的宇宙监督假想：它使留在黑洞外面的观察者不致受到发生在奇点处的可预见性崩溃的影响，但它对那位不幸落到黑洞里的可怜的航天员却是爱莫能助。

广义相对论方程存在一些解，我们的航天员在这些解中可能看到裸奇点：他也许能避免撞到奇点上去，相反地穿过一个"虫洞"来到宇宙的另一区域。看来这给在时空内的旅行提供了大的可能性。但是不幸的是，所有这些解似乎都是非常不稳定的；最小的干扰，譬如一个航天员的存在就会使之改变，以至于他还没能看到此奇点，就撞上去而终结了他的时间。换言之，奇点总发生在他的将来，而绝不会发生在他的过去。宇宙监督假想强的版本是说，在一个现实的解里，奇点总是要么整个存在于将来（如引力坍缩的奇点），要么整个存在于过去（如大爆炸）。我强烈地相信宇宙监督，这样我就和加州理工学院的基帕·索恩和约翰·普勒斯基尔打赌，认为它总是成立的。由于找到了一些解的例子，在非常远处可以看得见其奇点，所以我在技术的层面上输了。这样，我必须遵照协约还清赌债，也就是必须把他们的裸露遮盖住。但是我可以宣布道义上的胜利。这些裸奇点是不稳定的：最小的干扰就会导致这些奇点消失，或者躲到事件视界后面去。所以它们在实际情形下不会发生。

事件视界，也就是时空中不可逃逸区域的边界，其行为犹如围绕着黑洞的单向膜：物体，譬如粗心的航天员，能通过事件视界落到黑洞里去，但是没有任何东西可以通过事件视界而逃离黑洞。（记住事

件视界是企图逃离黑洞的光在时空中的路径，而且没有任何东西可以比光行进得更快）。人们可以将诗人但丁针对地狱入口所说的话恰到好处地应用于事件视界："从这里进去的人必须抛弃一切希望。"任何东西或任何人，一旦进入事件视界，就会很快地到达无限致密的区域和时间的终点。

广义相对论预言，运动的重物会导致引力波的辐射，那是以光的速度行进的空间曲率的涟漪。引力波和电磁场的涟漪光波相类似，但是要探测到它则困难得多。引力波引起邻近自由落体之间距离的非常微小的变化，由此可以观察到它。在美国、欧洲和日本正在建造一些检测器，将把十万亿亿（1后面跟21个0）分之一的位移，或者把在10英里距离中的比一个原子核还小的位移测量下来。

就像光一样，引力波带走了发射它们的物体的能量。因为任何运动中的能量都会被引力波的辐射带走，所以可以预料，一个大质量物体的系统最终会趋向于一种不变的状态。（这和扔一块软木到水中的情况相当类似：起先翻上翻下折腾了好一阵，但是随着涟漪将其能量带走，它最终平静下来。）例如，围绕着太阳公转的地球即产生引力波。其能量损失的效应就要改变地球的轨道，使之逐渐越来越接近太阳，最后撞到太阳上，归于一种不变的状态。在地球和太阳的情形下，能量损失率非常小——大约只能点燃一个小电热器。这意味着要用大约1000亿亿亿年地球才会撞到太阳上，没有必要立即为之担忧！地球轨道改变极其缓慢，根本观测不到。但几年以前，在称为PSR1913＋16（PSR表示"脉冲星"，一种特别的发射出射电波规则脉冲的中子星）的系统中观测到这同一效应。此系统由两个相互围绕着公转的中子星组成，由于引力波辐射，它们的能量损失，使它们相互沿着螺旋线轨道靠近。J·H·泰勒和R·A·荷尔西由于对广义相对论的这一证实获得1993年的诺贝尔奖。大约3亿年后它们将会碰撞。它们在碰撞之前，将会公转得这么快速，发射出的引力波，足以让像LIGO这样的检测器接收到。

在恒星引力坍缩形成黑洞时，运动会快得多，这样携带走能量的速率就会高得多。因此不用太长的时间就会达到不变的状态。这最终的状态将会是怎样的呢？人们会以为，它将依赖于形成黑洞的恒星的所有复杂特征——不仅它的质量和转动速度，而且恒星不同部分的不同密度以及恒星内气体的复杂运动。而如果黑洞就像坍缩形成它们的原先物体那样变化多端，那么一般来讲，对黑洞作任何预言都会非常困难。

然而，加拿大科学家威纳·伊斯雷尔（他生于柏林，在南非长大，在爱尔兰得到博士学位）在1967年使黑洞研究发生了彻底的改变。伊斯雷尔指出，根据广义相对论，非旋转的黑洞必须是非常简单的；它们是完美的球形，其大小只依赖于它们的质量，并且任何两个这样的同质量的黑洞必须等同。事实上，它们可以用爱因斯坦的特解来描述，这个解是在广义相对论发现后不久的1917年被卡尔·施瓦兹席尔德找到的。起初许多人，其中包括伊斯雷尔本人，认为，既然黑洞必须是完美的球形，一个黑洞只能由一个完美球形物体坍缩形成。因此，任何实际的恒星——从来都不是完美的球形——只会坍缩形成一个裸奇点。

然而，对于伊斯雷尔的结果，一些人，特别是罗杰·彭罗斯和约翰·惠勒提倡一种不同的解释。他们论证道，牵涉恒星坍缩的快速运动表明，其释放出来的引力波使之越来越接近于球形，到它终结于静态的时刻，就变成准确的球形。按照这种观点，任何非旋转恒星，不管其形状和内部结构如何复杂，在引力坍缩之后都将终结于一个完美的球形黑洞，其大小只依赖于它的质量。这种观点得到进一步计算的支持，并且很快就被大家接受。

伊斯雷尔的结果只处理了由非旋转物体形成的黑洞。1963年，新西兰人罗伊·克尔找到了广义相对论方程的描述旋转黑洞的一组解。这些“克尔”黑洞以恒常速度旋转，其大小与形状只依赖于它们的质量和旋转的速度。如果旋转为零，黑洞就是完美的球形，这解就和施瓦兹

席尔德解一样。如果旋转不为零，黑洞在赤道附近就会鼓出去（正如地球或太阳由于旋转而鼓出去一样），而旋转得越快则鼓得越厉害。由此人们猜测，如将伊斯雷尔的结果推广到包括旋转物体的情形，则任何旋转物体坍缩形成黑洞后，将最后终结于由克尔解描述的一个稳态。

1970年，我在剑桥的一位同事和研究生同学布兰登·卡特为证明此猜测跨出了第一步。他指出，假定一个稳态的旋转黑洞，正如一个自旋的陀螺那样，有一个对称轴，则它的大小和形状，只由它的质量和旋转速度决定。然后我在1971年证明了，任何稳态的旋转黑洞确实有这样的一个对称轴。最后在1973年，在伦敦国王学院任教的大卫·罗宾逊利用卡特和我的结果证明了这猜测是对的：这样的黑洞确实必须是克尔解。这样，在引力坍缩之后，一个黑洞必须最终演变成一种能够旋转，但是不能搏动的态。此外，它的大小和形状，只决定于它的质量和旋转速度，而与坍缩形成黑洞的原先物体的性质无关。此结果因如下一句格言而众所周知："黑洞没有毛。""无毛"定理具有巨大的实际重要性，因为它极大地限制了黑洞的可能类型。因此，人们可以制造可能包含黑洞的对象的详细模型，再将此模型的预言和观测相比较。因为在黑洞形成之后，我们所能测量的只是有关坍缩物体的质量和旋转速度，所以"无毛"定理还意味着，有关这物体的非常大量的信息，在黑洞形成时损失了。下一章我们将会理解这个意义。

黑洞是科学史上极为罕见的情形之一，在没有任何观测到的证据说明其理论是正确的情形下，作为数学的模型被发展到非常详尽的地步。的确，这经常是黑洞反对者的主要论据：人们怎么能相信这样的物体，其仅有的证据是基于令人怀疑的广义相对论的计算呢？然而，1963年，加利福尼亚的帕罗玛天文台的天文学家马丁·施密特测量了在称为3C273（即是剑桥射电源编目第三类的273号）射电源方向的一个黯淡的类星体的红移。他发现引力场不可能引起这么大的红移——如果它是引力红移，这类星体质量必须这么大，并且离我们必须

这么近，势必干扰太阳系中的行星轨道。这暗示这个红移是由宇宙的膨胀引起的，进而表明此物体离我们非常遥远。由于在这么远的距离还能观察到，它必须非常亮，也就是必须辐射出大量的能量。人们会想到，产生这么大能量的唯一机制看来不仅是一个恒星，而是一个星系的整个中心区域的引力坍缩。人们还发现了许多其他相似的类星体，它们都有很大的红移。但是它们都离开我们太远了，所以对之进行观察太困难了，不能给黑洞提供结论性的证据。

1967年，剑桥的一位研究生约瑟琳·贝尔发现了天空发射出射电波的规则脉冲的物体，这对黑洞存在的预言带来了进一步的鼓舞。起初贝尔和她的导师安东尼·赫维许以为，他们可能和我们星系中的外星文明进行了接触！我清楚地记得在宣布他们发现的讨论会上，他们将这四个最早发现的源称为LGM1-LGM4，LGM表示“小绿人”（“Little Green Man”）的意思。然而，最终他们和其他所有人都得到了不那么浪漫的结论，这些被称为脉冲星的物体，事实上是旋转的中子星。因为它们的磁场和周围物质复杂的相互作用，这些中子星发出射电波的脉冲。这对于写空间探险的作者而言是个坏消息，但对于我们这些当时相信黑洞的少数人来说，是非常大的希望——这是中子星存在的第一个正的证据。中子星的半径大约为10英里，只是恒星变成黑洞的临界半径的几倍。如果一颗恒星能坍缩到这么小的尺度，预料其他恒星能坍缩到更小的尺度而成为黑洞，就是理所当然的了。

按照黑洞定义，它不能发出光，我们何以希望能检测到它呢？这有点像在煤库里找黑猫。庆幸的是，有一种办法。正如约翰·米歇尔在他1783年的先驱性论文中指出的，黑洞仍然将它的引力作用到周围的物体上。天文学家观测了许多系统，在这些系统中，两颗恒星由于相互之间的引力吸引而相互围绕着运动。他们还观察到了这样的系统，其中只有一颗可见的恒星围绕着另一颗看不见的伴星运动。人们当然不能立即得出结论说，这伴星即为黑洞——它可能仅仅是一颗黯淡的看不见的恒星而已。然而，这种系统中的一些，像叫做天鹅X-1的

（图6.2）那样，也是强X射线源。对这现象的最好解释是，物质从可见星的表面被吹起来，当它落向不可见的伴星时，形成螺旋状运动（这和水从浴缸流出很相似），并且变得非常热，发出X射线（图6.3）。为了使这机制起作用，不可见物体必须非常小，像白矮星、中子星或黑洞那样。通过观测那颗可见星的轨道，人们可以确定不可见物体的最小的可能质量。在天鹅X-1的情形，这大约是太阳质量的6倍。按照昌德拉塞卡的结果，它的质量太大了，既不可能是白矮星，也不可能是中子星。因此，看来它只能是一个黑洞。



图6.2　在靠近照片中心的两个恒星之中最亮的那颗是天鹅X-1,被认为是由互相绕着旋转的一个黑洞和一个正常恒星组成

还有其他不包含黑洞的解释天鹅X-1的模型，但是所有这些都相当牵强附会。黑洞看来是对该观测的仅有的真正自然的解释。尽管如此，我和加州理工学院的基帕·索恩打赌说，天鹅X-1不包含一个黑洞！这对我而言是一种保险的形式。我对黑洞作了许多研究，如果发现黑洞不存在，而这一切都成为徒劳。但在这种情形下，我将得到赢得打赌的安慰，他要给我订阅4年的《私家侦探》杂志。事实上，从

我们打赌的1975年迄今，虽然天鹅X-1的情形并没有改变太多，但是人们已经积累了这么多对黑洞有利的其他观测证据，我只好认输。我进行了约定的赔偿，那就是给索恩订阅一年的《藏春阁》，这使他开放的妻子相当恼火。

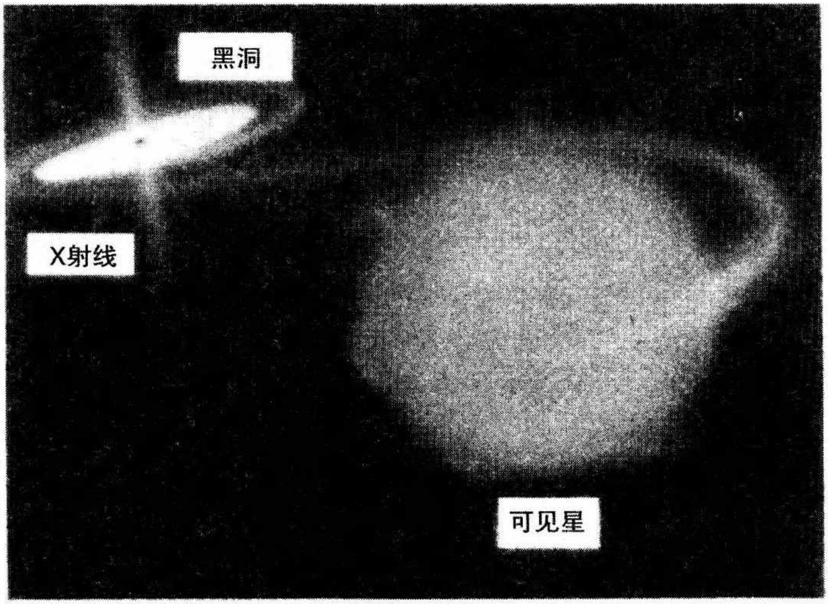



图6.3

现在，在像我们的星系和两个名叫麦哲伦星云的邻近星系的系统中，我们还有几个类似天鹅X-1的黑洞的证据。然而，几乎可以肯定，黑洞的数量比这多得太多了！在宇宙的漫长历史中，很多恒星肯定烧尽了它们的核燃料并坍缩了。黑洞的数目甚至比可见恒星的数目要大得多。仅仅在我们的星系中，大约总共有1000亿颗可见恒星。这样巨大数量的黑洞的额外引力就能解释为何目前我们的星系以现有的速率转动：仅用可见恒星的质量是不足以说明这一点的。我们还有某些证据表明，在我们星系的中心有一个大得多的黑洞，其质量大约是太阳的10万倍。星系中的恒星若十分靠近这个黑洞时，作用在它的近端和远端上的引力之差或潮汐力会将其撕开。它们的遗骸以及摆脱其

他恒星的气体将落到黑洞上去。正如在天鹅X-1的情形那样，气体将以螺旋形轨道向里运动，并且被加热，虽然没有到那种程度。它没有热到足以发出X射线，但是它可以用来说明在星系中心观测到的非常致密的射电波和红外线源。

人们认为，在类星体的中心是类似的，但质量更大的黑洞，其质量大约为太阳的1亿倍。例如，用哈勃望远镜对称为M87的星系进行的观测揭示出，它含有直径130光年的气体盘，该盘围绕着20亿倍太阳质量的中心物体旋转。这只能是一个黑洞。只有落入此超重的黑洞的物质才能提供足够强大的能源，用以解释这些物体释放出的巨大能量。当物质旋入黑洞，它将使黑洞往同一方向旋转，使黑洞产生一个磁场，这个磁场和地球的磁场颇为相像。落入的物质会在黑洞附近产生能量非常高的粒子。该磁场是如此之强，能将这些粒子聚焦成沿着黑洞旋转轴，也即在它的北极和南极方向往外喷射的射流。在许多星系和类星体中确实观察到这类射流。人们还可以考虑存在质量比太阳质量小很多的黑洞的可能性。因为它们的质量比昌德拉塞卡极限低，所以不能由引力坍缩产生：这样小质量的恒星，甚至在耗尽了自己的核燃料之后，还能支持自己对抗引力。只有当物质由非常巨大的外界压力压缩成极端紧密的状态时，才能形成小质量的黑洞。一个巨大的氢弹可提供这样的条件：物理学家约翰·惠勒曾经计算过，如果将世界海洋里所有的重水制成一个氢弹，则它可以将中心的物质压缩到产生一个黑洞。（当然，那时没有一个人能残留下来观察它！）比较实在的一种可能性是：在极早期宇宙的高温和高压条件下可能产生这样小质量的黑洞。因为只有一个比平均值更紧密的小区域，才能以这样的方式被压缩形成一个黑洞，所以只有当早期宇宙不是完全光滑的和均匀时，这才有可能形成黑洞。但是我们知道，早期宇宙一定存在一些无规性，否则现在宇宙中的物质分布仍然会是完全均匀的，而不能结块形成恒星和星系。

很清楚，为了说明恒星和星系的无规性是否导致形成相当数目

的“太初”黑洞，依赖于早期宇宙中条件的细节。这样，如果我们能够确定现在有多少太初黑洞，我们就能对宇宙的极早期阶段了解很多。质量大于10亿吨（一座大山的质量）的太初黑洞，只能通过它们对其他可见物质或宇宙膨胀的影响被探测到。然而，正如我们将要在下一章看到的，黑洞毕竟不是真黑：它们像一个热体一样发热发光，它们越小则发热发光得越厉害。所以，看起来荒谬，而事实上却是，也许小的黑洞可以比大的黑洞更容易探测到！

# 第七章 黑洞不是这么黑的

在1970年以前，我关于广义相对论的研究，主要集中于是否存在一个大爆炸奇点。然而，同年11月我的女儿露西出生后不久的一个晚上，当我上床时，开始思考黑洞的问题。我的残废使得这个过程相当缓慢，这样我有大量时间。那时候还不存在关于时空的那些点是在黑洞之内还是在黑洞之外的准确定义。我已经和罗杰·彭罗斯讨论过将黑洞定义为不能逃逸到远处的事件集合的想法，这也就是现在被广泛接受的定义。它意味着，黑洞边界——即事件视界——是由刚好不能从黑洞逃逸，而只在边缘上永远盘旋的光线在时空里的路径形成的（图7.1）。这有点像从警察那里逃开，但是仅仅维持比警察快一步，而不能彻底逃脱的情景！

我忽然意识到，这些光线的路径永远不可能相互靠近。如果它们靠近，它们最终就必定相撞。这正如和另一个往相反方向逃离警察的人相遇一样——你们俩都会被抓住（或者，在这种情形下落到黑洞中去）。但是，如果这些光线被黑洞吞没，那它们就从未在黑洞的边界上呆过。所以在事件视界上的光线的路径必须永远相互平行运动或相互散开。另一种看到这一点的方法是，事件视界，亦即黑洞边界，正像一个影子的边缘——一个即将临头的灾难的影子。如果你看到在远距离上的一个源，譬如太阳，投下的影子，就能明白边缘上的光线不会相互靠近。

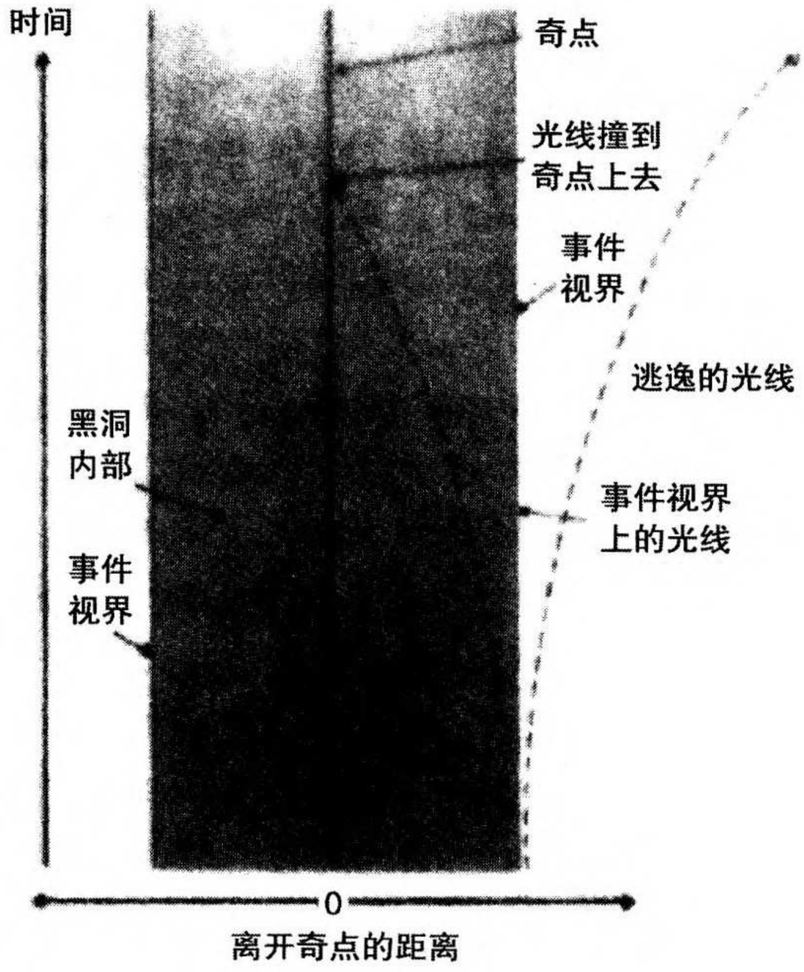



图7.1

如果从事件视界（亦即黑洞边界）来的光线永不相互靠近，则事件视界的面积可以保持不变或者随时间增大，但它永远不会减小——因为这意味着至少边界上的一些光线必须互相靠近。事实上，每当物质或辐射落到黑洞中去，这面积就会增大（图7.2）；或者如果两个黑洞碰撞并合并成一个单独的黑洞，这最后的黑洞的事件视界面积就会大于或等于原先黑洞事件视界面积的总和（图7.3）。事件视界面

积的非减性质给黑洞的可能行为加上了重要的限制。我为我的发现如此激动，以至于当夜没睡多少。第二天，我给罗杰·彭罗斯打电话，他同意我的结果。我想，事实上他此前已经意识到了这个面积的性质。然而，他使用了稍微不同的黑洞定义。他没有意识到，假定黑洞已经终止于不随时间变化的状态，按照这两种定义，黑洞的边界并因此其面积都应是一样的。

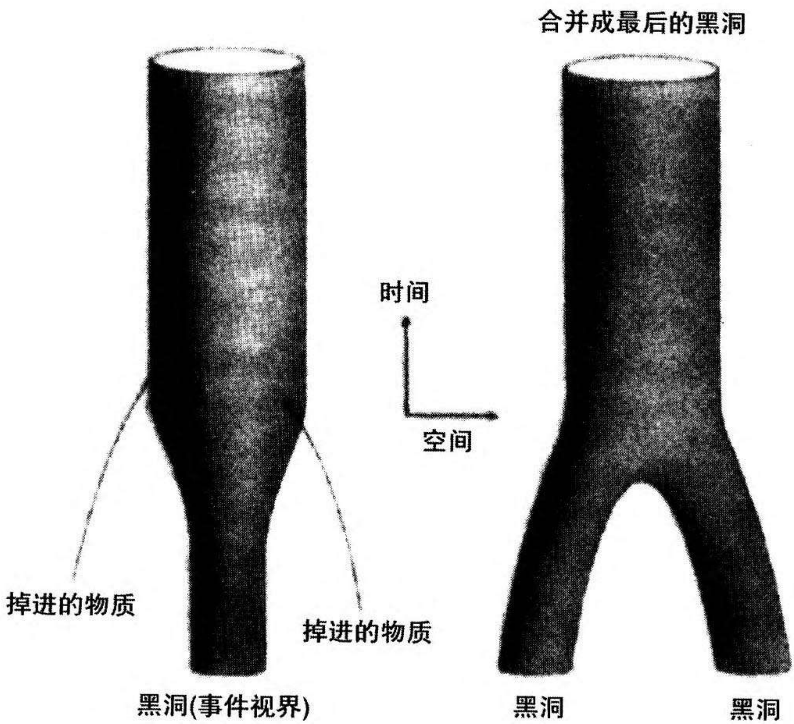



图7.2（左），图7.3（右）

人们非常容易从黑洞面积的非减行为联想起被叫做熵的物理量的行为。熵是测量一个系统的无序的程度。常识告诉我们，如果不进行外部干涉，事物总是倾向于增加它的无序度。（你只要停止保养房子就会看到这一点！）人们可以从无序中创造出有序来（例如你可以油漆房子），但是必须消耗精力或能量，这样减少了可利用的有序能量

的数量。

热力学第二定律是这个观念的一个准确描述。它陈述道：一个孤立系统的熵总是增加的，并且将两个系统连接在一起时，其合并系统的熵大于所有单独系统熵的总和。譬如，考虑一盒气体分子的系统。分子可以认为是不断相互碰撞，并不断从盒子壁反弹回来的康乐球。气体的温度越高，分子运动得越快，这样它们撞击盒壁越频繁也越厉害，而且它们作用到壁上的向外的压力越大。假定初始时所有分子被一隔板限制在盒子的左半部。如果接着将隔板除去，这些分子将趋向散开并充满盒子的两半。在以后的某一时刻，所有这些分子偶尔会都呆在右半部或回到左半部，但占绝对优势的可能性是，分子的数目在左右两半大致相同。这种状态比原先的所有分子都在一个半部的状态更加无序。因此，人们说气体的熵增加了。类似地，假定我们从两个盒子开始，将一个盒子充满氧分子，另一个盒子充满氮分子。如果把两个盒子连在一起并移去中间的壁，则氧分子和氮分子就开始混合。在后来的时刻，最可能的状态是两个盒子都充满了相当均匀的氧分子和氮分子的混合物。这种状态比原先分开的两盒的初始状态更无序，即具有更大的熵。

和其他科学定律，譬如牛顿引力定律相比，热力学第二定律的状况相当不同。例如，它只是在绝大多数的而非所有情形下成立。在以后某一时刻，我们第一个盒子中的所有气体分子在盒子的一半被发现的概率只有几万亿分之一，但它们可能发生。然而，如果附近有一黑洞，似乎存在一种非常容易的方法违反第二定律：只要将一些具有大量熵的物体，譬如一盒气体，抛进黑洞里。黑洞之外物体的总熵就会减少。当然，人们仍然可以说，包括黑洞里的熵的总熵没有降低——但是由于没有办法看到黑洞里面，我们不能知道里面物体的熵为多少。如果黑洞具有某一特征，黑洞外的观察者因之可知道它的熵，并且只要携带熵的物体一落入黑洞，它就会增加，那将是很美妙的。紧接着上述的黑洞面积定理的发现，即只要物体落入黑洞，它的事件视

界面积就会增加，普林斯顿大学一位名叫雅可布·柏肯斯坦的研究生提出，事件视界的面积即是黑洞熵的量度。由于携带熵的物质落到黑洞中时，它的事件视界的面积会增加，这样就使黑洞外物质的熵和事件视界面积的和永远不会降低。

看来在大多数情况下，这个建议防止热力学第二定律受到违背。然而还有一个致命的瑕疵。如果一个黑洞具有熵，那它也应该有温度。但具有特定温度的物体必须以一定的速率发出辐射。从日常经验知道：只要将火钳在火上加热，它就会发光发热，发出辐射。但在低温下物体也发出辐射；只是因为辐射量相当小，在通常情况下没有注意到。为了防止违反热力学第二定律，这辐射是必需的。所以黑洞必须发出辐射。但正是按照其定义，黑洞被认为是不发出任何东西的物体。因此，黑洞的事件视界的面积似乎不能认为是它的熵。1972年，我和布兰登·卡特以及美国同事詹姆·巴丁合写了一篇论文，在论文中我们指出，虽然在熵和事件视界的面积之间存在许多相似点，但还存在着这个致命的困难。我必须承认，写此文章的部分动机是因为被柏肯斯坦激怒，我觉得他滥用了我的事件视界面积增加的发现。然而，最后发现，他基本上还是正确的，虽然是在一种他肯定没有预料到的情形下。

1973年9月我访问莫斯科时，和苏联两位最主要的专家雅可夫·捷尔多维奇和亚历山大·斯塔拉宾斯基讨论黑洞问题。他们说服我，按照量子力学不确定性原理，旋转黑洞应该产生并辐射粒子。在物理学的基础上，我相信他们的论点，但是不喜欢他们计算辐射所用的数学方法。因此，我着手设计一种更好的数学处理方法，并于1973年11月底在牛津的一次非正式讨论会上将其公布于众。那时我还没计算出实际上有多少辐射。我预料要发现的正是捷尔多维奇和斯塔拉宾斯基预言的从旋转黑洞发出的辐射。然而，当我做了计算，使我既惊奇又恼火的是，我发现甚至非旋转黑洞显然也应以不变速率产生和发射粒子。起初我以为这种辐射表明我使用的一种近似无效。我担心如果柏

肯斯坦发现了这个情况，他就一定会用它去进一步支持他关于黑洞熵的思想，而我仍然不喜欢这种思想。然而，我越仔细推敲，越觉得这近似其实应该有效。但是，最后使我信服这辐射是真实的理由是，这辐射的粒子谱刚好是一个热体辐射的谱，而且黑洞以刚好防止第二定律被违反的正确速率发射粒子。此后，其他人用多种不同的形式重复了这个计算。他们所有人都证实了黑洞必须如同一个热体那样发射粒子和辐射，其温度只依赖于黑洞的质量——质量越大则温度越低。

我们知道，任何东西都不能从黑洞的事件视界之内逃逸出来，黑洞怎么可能发射粒子呢？量子理论给我们的回答是，粒子不是从黑洞里面出来的，而是从紧靠黑洞的事件视界的外面的“空虚的”空间来的！我们可以用以下的方法去理解这个：我们以为是“空虚的”空间不能是完全空的，因为那就意味着诸如引力场和电磁场的所有场都必须刚好是零。然而场的数值和它的时间变化率如同粒子的位置和速度那样：不确定性原理意味着，人们对其中的一个量知道得越准确，则对另一个量知道得越不准确。所以在空虚的空间里场不可能严格地被固定为零，因为那样它就既有准确的值（零）又有准确的变化率（也是零）。场的值必须有一定的最小的不确定性量或量子起伏。人们可以将这些起伏理解为光或引力的粒子对，它们在某一时刻同时出现，相互离开，然后又相互靠近，而且相互湮灭。这些粒子正如同携带太阳引力的虚粒子：它们不像真的粒子那样，能用粒子探测器直接观察到。然而，它们的间接效应，例如原子中的电子轨道能量发生的微小变化，可被测量出，并和理论预言一致的程度，令人十分惊讶。不确定性原理还预言了存在类似的虚的物质粒子对，例如电子对和夸克对。然而在这种情形下，粒子对的一个成员为粒子，而另一成员为反粒子（光和引力的反粒子和粒子相同）。

因为能量不能无中生有，所以粒子反粒子对中的一个伴侣具有正能量，而另一个具有负能量。由于在正常情况下实粒子总是具有正能量，所以具有负能量的那一个粒子注定是短命的虚粒子。因此，它必

须找到它的伴侣并与之相互湮灭。然而，因为实粒子要花费能量抵抗大质量物体的引力吸引才能将其推到远处，一颗实粒子的能量在接近大质量物体时比在远离时更小。正常情况下，这粒子的能量仍然是正的。但是黑洞里的引力是如此之强，甚至在那里实粒子的能量都可以是负的。因此，如果存在黑洞，带有负能量的虚粒子落到黑洞里可能变成实粒子或实反粒子。这种情形下，它不再需要和它的伴侣相互湮灭了。它被抛弃的伴侣也可以落到黑洞中去。或者由于它具有正能量，也可以作为实粒子或实反粒子从黑洞的邻近逃走（图7.4）。对于一个远处的观察者而言，它就显得是从黑洞发射出来的粒子一样。黑洞越小，负能粒子在变成实粒子之前必须走的距离越短，这样黑洞发射率和表观温度也就越大。

辐射出去的正能量会被落入黑洞的负能粒子流平衡。按照爱因斯坦方程$E=mc^2$（E是能量，m是质量，c为光速），能量和质量成正比。因此，往黑洞去的负能量流减小它的质量。随着黑洞损失质量，它的事件视界面积变得更小，但是它发射出的辐射的熵过量地补偿了黑洞的熵的减少，所以第二定律从未被违反过。

还有，黑洞的质量越小，其温度就越高。这样，随着黑洞损失质量，它的温度和发射率增加，因而它的质量损失得更快。当黑洞的质量最后变得极小时会发生什么，人们并不很清楚。但是最合理的猜想是，它最终将会在一次巨大的，相当于几百万颗氢弹爆炸的辐射暴中消失殆尽。

一个具有几倍太阳质量的黑洞只具有一千万分之一度的绝对温度。这比充满宇宙的微波辐射的温度（大约2.7K）要低得多，所以这种黑洞的辐射比它吸收的还要少。如果宇宙注定继续永远膨胀下去，微波辐射的温度就会最终减小到比这黑洞的温度还低，它就开始损失质量。但是即使到了那时候，它的温度是如此之低，以至于要用100亿亿亿亿亿亿亿年（1后面跟66个0）才全部蒸发完。这比宇宙的年龄长得多了，宇宙的年龄大约只有100至200亿年（1或2后面跟10个

0）。另一方面，正如第六章提及的，在宇宙的极早期阶段存在由于无规性引起的坍缩而形成的质量极小的太初黑洞。这样的小黑洞会有高得多的温度，并以大得多的速率发出辐射。具有10亿吨初始质量的太初黑洞的寿命大体和宇宙的年龄相同。初始质量比这小的太初黑洞应该已蒸发完毕，但那些比这稍大的黑洞仍在辐射出X射线以及伽马射线。这些X射线和伽马射线像光波，只是波长短得多。这样的黑洞几乎不配这黑的绰号：它们实际上是白热的，正以大约1万兆瓦的功率发射能量。

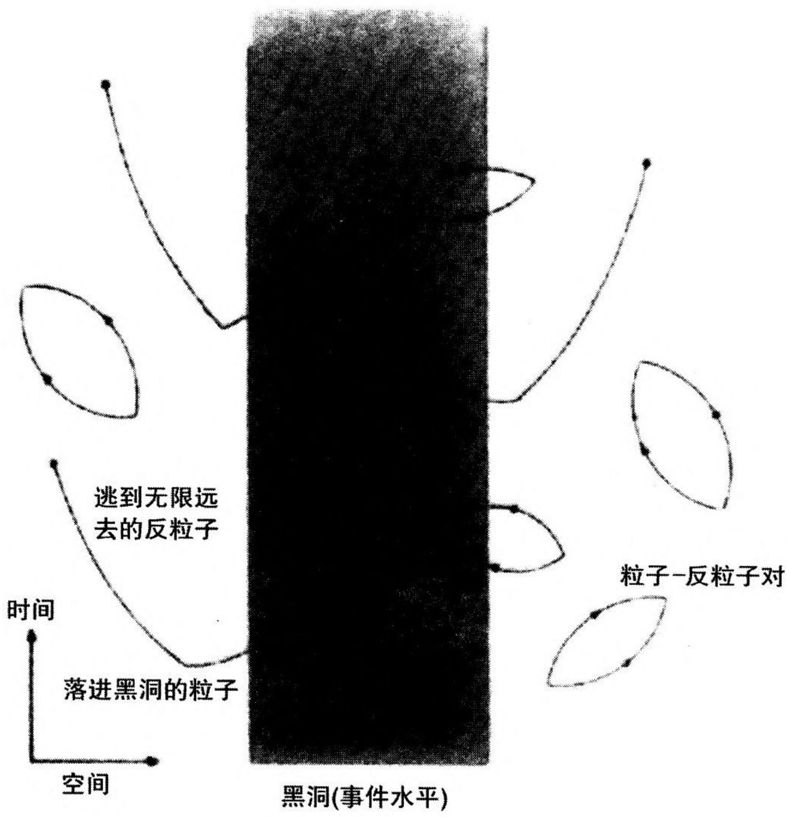



图7.4

一个这样的黑洞可以开动10个大型的发电站，只要我们能够驾驭

黑洞的功率就好了。然而，这是非常困难的：这黑洞把和一座山差不多的质量压缩成比万亿分之一英寸，亦即一个原子核的尺度还小！如果你在地球表面上有这样的一个黑洞，就无法阻止它透过地面落到地球的中心。它会穿过地球而来回振动，直到最后停在地球的中心。所以仅有的放置黑洞并利用之发射出能量的地方是围绕着地球的轨道，而仅有的使它围绕地球公转的办法是，用在它之前的一个大质量的吸引力去拖它，这和在驴子前面放一根胡萝卜颇为相像。至少在最近的将来，这个设想并不现实。

但是，即使我们不能驾驭来自这些太初黑洞的辐射，我们观测到它们的机遇又如何呢？我们可以寻找太初黑洞在其主要生存期里发出的伽马射线辐射。虽然大部分黑洞在很远以外的地方，从它们来的辐射非常弱，但是从它们全体来的总辐射是可以检测得到的。我们确实观察到这样的一个伽马射线背景：图7.5表示观察到的强度随频率（每秒波动的次数）的变化。然而，这个背景可以，并且大概是由除了太初黑洞以外的过程产生的。图7.5中的点线指出，如果每立方光年平均有300个太初黑洞，它们所发射的伽马射线的强度应如何随频率变化。因此可以说，伽马射线背景的观测并没给太初黑洞提供任何肯定的证据。但它们明确告诉我们，在宇宙中平均每立方光年不可能有多于300个太初黑洞。这个极限表明，太初黑洞最多只能构成宇宙中一百万分之一的物质。

由于太初黑洞是如此稀罕，似乎不太可能存在一个近到我们可以将其当作一个单独的伽马射线源来观察的黑洞。但是由于引力会将太初黑洞往任何物体处拉近，所以它们在星系里面和附近应该会更稠密得多。虽然伽马射线背景告诉我们，平均每立方光年不可能有多于300个太初黑洞，但它并没有告诉我们，太初黑洞在我们星系中有多么普遍。譬如讲，如果它们的密度比这个普遍100万倍，则离开我们最近的黑洞可能大约在10亿千米远，或者大约是已知的最远的行星——冥王星那么远。在这个距离上去探测黑洞恒定的辐射，即使其功

率为1万兆瓦，仍是非常困难的。为了观测到一个太初黑洞，人们必须在合理的时间间隔里，譬如一星期，从同方向检测到几个伽马射线量子。否则，它们仅可能是背景的一部分。因为伽马射线有非常高的频率，从普朗克量子原理得知，每一伽马射线量子都具有非常高的能量，这样甚至辐射1万兆瓦都不需要许多量子。而要观测到从冥王星这么远来的这些稀少的粒子，需要一个比任何迄今已经建造的更大的伽马射线探测器。况且，由于伽马射线不能穿透大气层，此探测器必须放置到太空。

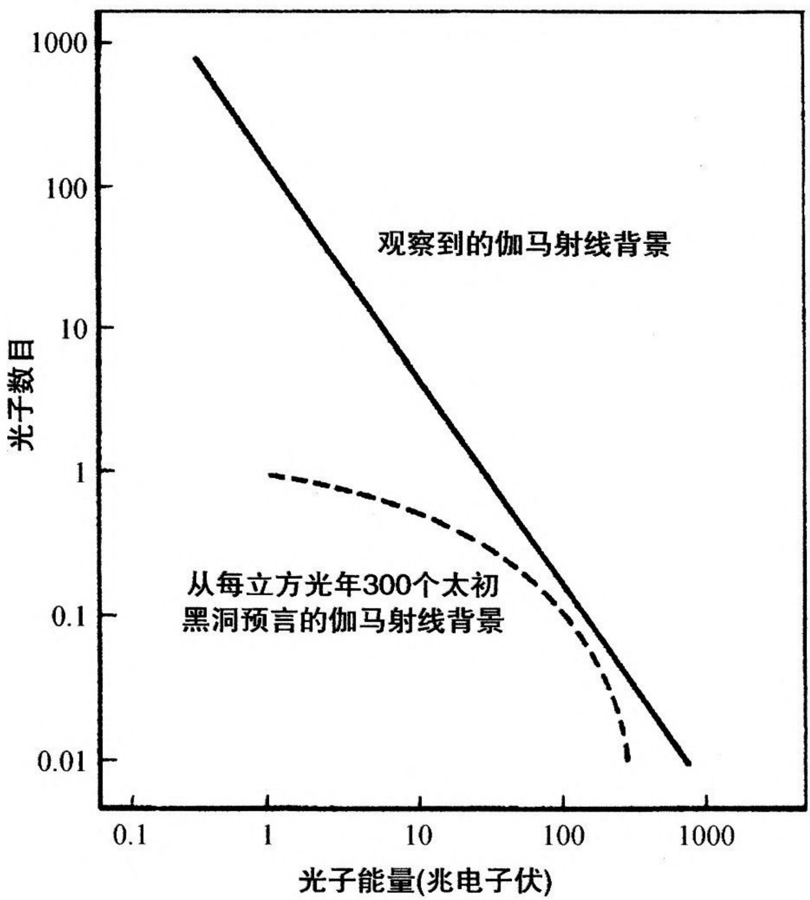



图7.5

当然，如果一颗像冥王星这么近的黑洞已达到它生命的末期并要爆炸开来，很容易检测其最后辐射暴。但是，如果一个黑洞已经发射了100至200亿年，不在过去或将来的几百万年里，而是在未来的若干年里到达它生命终点的可能性真是微不足道！所以在你的研究津贴用光之前，为了有一合理的机会看到爆炸，必须找到在大约1光年距离之内检测任何爆炸的方法。事实上，原先建造来监督违反禁止核试验条约的卫星检测到了从太空来的伽马射线暴。这些每个月似乎发生16次左右，并且大体均匀地分布在天空的所有方向上。这表明它们起源于太阳系之外，否则的话，我们可以预料它们要集中于行星轨道面上。这种均匀分布还表明，这些伽马射线源要么处于银河系中离我们相当近的地方，要么就在它的外围的宇宙学距离之处，因为否则的话，它们又会集中于星系的平面附近。在后者的情形下，产生伽马射线暴所需的能量实在太大，微小的黑洞根本提供不起。但是如果这些源以星系的尺度衡量和我们邻近，那就可能是正在爆发的黑洞。我非常希望这种情形成真，但是我必须承认，还可以用其他方式来解释伽马射线暴，例如中子星的碰撞。未来几年的新观测，尤其是像LIGO这样的引力波探测器，应该能使我们发现伽马射线暴的起源。

即使对太初黑洞的寻求证明是否定的，看来可能会是这样，仍然给了我们关于极早期宇宙的重要信息。如果早期宇宙曾经是混沌或不规则的，或者如果物质的压力曾经很低，可以预料到会产生比我们由观测伽马射线背景设下的极限更多得多的太初黑洞。只有当早期宇宙是非常光滑和均匀的，并有很高的压力，人们才能解释为何没有可观数目的太初黑洞。

黑洞辐射的思想是这种预言的第一例，它以基本的方式依赖于本世纪两个伟大理论，即广义相对论和量子力学。因为它推翻了已有的观点，所以一开始就引起了许多反对：“黑洞怎么能辐射东西？”当我在牛津附近的卢瑟福一阿普顿实验室的一次会议上，第一次宣布我的计算结果时，受到了普遍质疑。我讲演结束后，会议主席伦敦国王学

院的约翰·泰勒宣布这一切都是毫无意义的。他甚至为此还写了一篇论文。然而，最终包括约翰·泰勒在内的大部分人都得出结论：如果我们关于广义相对论和量子力学的其他观念是正确的，那么黑洞必须像热体那样辐射。这样，即使我们还不能找到一个太初黑洞，大家相当普遍地同意，如果找到的话，它必须正在发射出大量的伽马射线和X射线。

黑洞辐射的存在似乎意味着，引力坍缩不像我们曾经认为的那样是最终的、不可逆转的。如果一个航天员落到黑洞中去，黑洞的质量将增加，但是最终这额外质量的等效能量将会以辐射的形式回到宇宙中去。这样，此航天员在某种意义上被“再循环”了。然而，这是一种非常可怜的不朽，因为当航天员在黑洞里被撕开时，他的任何个人的时间的概念几乎肯定都达到了终点！甚至最终从黑洞辐射出来的粒子的种类，一般来说都和构成这航天员的不同：这航天员所遗留下来的仅有特征是他的质量或能量。

当黑洞的质量大于几分之一克时，我用以推导黑洞辐射的近似应是很有效的。但是，当黑洞在它的生命晚期，质量变成非常小时，这近似就失效了。最可能的结果看来是，它至少从宇宙的我们这一区域消失了，带走了航天员和可能在它里面的任何奇点（如果其中确有一个奇点的话）。这是量子力学能够去掉广义相对论预言的奇点的第一个迹象。然而，我和其他人在1974年使用的方法不能回答诸如在量子引力论中是否会发生奇性的问题。因此，从1975年以来，根据理查德·费恩曼对于历史求和的思想，我开始推导一种更强有力的量子引力论方法。这种方法对宇宙以及其诸如航天员之类的内容的开端和终结给出的答案，将在以下两章叙述。我们将会看到，虽然不确定性原理对于我们所有的预言的准确性都加上了限制，同时它却可以排除掉发生在时空奇点处的基本的不可预言性。

第八章

# 宇宙的起源和命运

从爱因斯坦广义相对论本身就能预言：时空在大爆炸奇点处开始，并会在大挤压奇点处（如果整个宇宙坍缩的话）或在黑洞中的一个奇点处（如果一个局部区域，譬如恒星坍缩的话）结束。任何落进黑洞的东西都会在奇点处毁灭，在外面只能继续感觉到它的质量的引力效应。另一方面，当考虑量子效应时，物体的质量和能量似乎会最终回到宇宙的其余部分，黑洞和在它当中的任何奇点会一道蒸发掉并最终消失。量子力学对大爆炸和大挤压奇点也能有同等戏剧性的效应吗？在宇宙的极早或极晚期，当引力场如此之强，量子效应不能不考虑时，究竟会发生什么？宇宙究竟是否有一个开端或终结？如果有的话，它们是什么样子的？

我在整个70年代主要研究黑洞，但在1981年参加在梵蒂冈由耶稣会组织的宇宙学会议时，我对于宇宙的起源和命运问题的兴趣被重新唤起。当天主教会试图对科学的问题发号施令，并宣布太阳围绕着地球运动时，对伽利略犯下了严重的错误。几个世纪后的现在，它决定邀请一些专家做宇宙学问题的顾问。在会议的尾声，教皇接见所有与会者。他告诉我们，在大爆炸之后的宇宙演化是可以研究的，但是我们不应该去过问大爆炸本身，因为那是创生的时刻，因而只能是上帝的事务。我心中窃喜，看来他并不知道，我刚在会议上作过的演讲的主题——时空有限而无界的可能性，这意味着它没有开端、没有创生的时刻。我不想去分享伽利略的厄运。我对伽利略之所以有一种强烈的认同感，其部分原因是我刚好出生于他死后的300年！

为了解释我和其他人关于量子力学如何影响宇宙的起源和命运的思想，必须首先按照所谓的“热大爆炸模型”来理解被广泛接受的宇宙历史。这是假定从早到大爆炸时刻起宇宙就可用弗里德曼模型来描述。在此模型中，人们发现当宇宙膨胀时，其中的任何物体或辐射都变得更凉（当宇宙的尺度大到2倍，它的温度就降低到一半。）由于温度即是粒子的平均能量——或速度的测度，宇宙的变凉对于其中的物质就会有较大的效应。在非常高的温度下，粒子能够运动得如此之快，可以逃脱任何由核力或电磁力将它们吸引在一起的作用。但是可以预料到，随着它们冷却下来，粒子相互吸引并且开始结块。更有甚者，连存在于宇宙中的粒子种类也依赖于温度。在足够高的温度下，粒子的能量是如此之高，只要它们碰撞就会产生很多不同的粒子/反粒子对一一并且，虽然其中一些粒子打到反粒子上去时会湮灭，但是它们产生得比湮灭得更快。然而，在更低的温度下，碰撞粒子具有较小的能量，粒子/反粒子对产生得不快——而湮灭则变得比产生更快。

就在大爆炸时，宇宙体积被认为是零，所以是无限热。但是，辐射的温度随着宇宙的膨胀而降低。大爆炸后的1秒钟，温度降低到约为100亿度，这大约是太阳中心温度的1000倍，亦即氢弹爆炸达到的温度。此刻宇宙主要包含光子、电子和中微子（极轻的粒子，它只受弱力和引力的作用）和它们的反粒子，还有一些质子和中子。随着宇宙的继续膨胀，温度继续降低，电子/反电子对在碰撞中的产生率就落到它们的湮灭率之下。这样，大多数电子和反电子相互湮灭掉了，产生出更多的光子，只剩下很少的电子。然而，中微子和反中微子并没有相互湮灭掉，因为这些粒子和它们自己以及其他粒子的作用非常微弱。这样，直到今天它们应该仍然存在。如果我们能观测到它们，就会为非常热的早期宇宙阶段的图象提供一个很好的检验。可惜现在它们的能量太低了，使得我们不能直接观察到。然而，如果中微子不是零质量，而是像近年的一些实验暗示的，自身具有小的质量，我们则可能间接地探测到它们：正如前面提到的那样，它们可以是“暗物

质”的一种形式，具有足够的引力吸引去遏止宇宙的膨胀，并使之重新坍缩。

在大爆炸后的大约100秒，温度降到了10亿度，也即最热的恒星内部的温度。在此温度下，质子和中子不再有足够的能量逃脱强核力的吸引，所以开始结合产生氘（重氢）的原子核。氘核包含一个质子和一个中子。然后，氘核和更多的质子、中子相结合形成氦核，它包含两个质子和两个中子，还产生了少量的两种更重的元素锂和铍。可以计算出，在热大爆炸模型中大约1/4的质子和中子变成了氦核，还有少量的重氢和其他元素。余下的中子会衰变成质子，这正是通常氢原子的核。

1948年，科学家乔治·伽莫夫和他的学生拉夫·阿尔法在一篇著名的合作的论文中，第一次提出了宇宙的热的早期阶段的图象。伽莫夫颇为幽默——他说服了核物理学家汉斯·贝特将他的名字加到这论文上面，使得列名作者为“阿尔法、贝特、伽莫夫”，正如最前面三个希腊字母：阿尔法、贝他、伽马：这特别适合于一篇关于宇宙开初的论文！他们在此论文中作出了一个惊人的预言：宇宙的热的早期阶段的辐射（以光子的形式）今天还应该在周围存在，但是其温度已被降低到只比绝对零度（-273℃）高几度。这正是彭齐亚斯和威尔逊在1965年发现的辐射。在阿尔法、贝特和伽莫夫写此论文时，对于质子和中子的核反应了解得不多，所以对于早期宇宙不同元素比例所作的预言相当不准确；但是，在用更好的知识重新进行这些计算之后，现在的结果已和我们的观测符合得非常好。况且，在解释宇宙为何应该有这么多氦时，用任何其他方法都是非常困难的。所以，我们相当确信，至少一直回溯到大爆炸后大约1秒钟为止，这个图象是正确无误的。

大爆炸后的几个钟头之内，氦和其他元素的产生就停止了。之后的100万年左右，宇宙仅仅是继续膨胀，没有发生什么事。最后，一旦温度降低到几千度，电子和核子不再有足够能量去战胜它们之间的电磁吸引力，就开始结合形成原子。宇宙作为整体，继续膨胀变冷，

但在一个比平均稍微密集些的区域，膨胀就会由于额外的引力吸引而缓慢下来。在一些区域膨胀最终会停止并开始坍缩。当它们坍缩时，在这些区域外的物体的引力拉力使它们开始很慢地旋转；当坍缩的区域变得更小，它会自转得更快——正如在冰上自转的滑冰者，缩回手臂时会自转得更快。最终，当区域变得足够小，它自转得快到足以平衡引力的吸引，碟状的旋转星系就以这种方式诞生了。另外一些区域刚好没有得到旋转，就形成了叫做椭圆星系的椭球状物体。这些区域之所以停止坍缩，是因为星系的个别部分稳定地围绕着它的中心公转，但星系整体并没有旋转。

随着时间流逝，星系中的氢和氦气体被分割成更小的星云，它们在自身引力下坍缩。当它们收缩时，其中的原子相互碰撞，气体温度升高，直到最后，热得足以开始热聚变反应。这些反应将更多的氢转变成氦，释放出的热增加了压力，因此使星云不再继续收缩。它们会稳定地在这种状态下，作为像太阳一样的恒星停留一段很长的时间，它们将氢燃烧成氦，并将得到的能量以热和光的形式辐射出来。质量更大的恒星需要变得更热，以平衡它们更强的引力吸引，使得其核聚变反应进行得极快，以至于它们在1亿年这么短的时间里将氢耗光。然后，它们会稍微收缩一点，而随着它们进一步变热，就开始将氦转变成像碳和氧这样更重的元素。但是，这一过程没有释放出太多的能量，所以正如在黑洞那一章描述的，危机就会发生了。人们不完全清楚下一步还会发生什么，但是看来恒星的中心区域很可能坍缩成一个非常致密的状态，譬如中子星或黑洞。恒星的外部区域有时会在称为超新星的巨大爆发中吹出来，这种爆发使星系中的所有恒星在相形之下显得黯淡无光。恒星接近生命终点时产生的一些重元素就被抛回到星系里的气体中去，为下一代恒星提供一些原料。因为我们的太阳是第二代或第三代恒星，是大约50亿年前由包含有更早超新星碎片的旋转气体云形成的，所以大约包含2%这样的重元素。云里的大部分气体形成了太阳或者喷到外面去，但是少量的重元素聚集在一起，形成了像地球这样的，现在作为行星围绕太阳公转的物体。

地球原先是非常热的，并且没有大气。在时间的长河中它冷却下来，并从岩石中散发气体得到了大气。我们无法在这早先的大气中存活。因为它不包含氧气，反而包含很多对我们有毒的气体，如硫化氢（即是使臭鸡蛋难闻的气体）。然而，存在其他能在这种条件下繁衍的原始的生命形式。人们认为，它们可能是作为原子的偶然结合，形成叫做宏观分子的大结构的结果，而在海洋中发展，这种结构能够将海洋中的其他原子聚集成类似的结构。它们就这样复制自己并繁殖。在有些情况下复制有些误差。这些误差通常使新的宏观分子不能复制自己，并最终被消灭。然而，一些误差会产生出新的宏观分子，它们会更有效地复制自己。因此它们具有优势，并趋向于取代原先的宏观分子。进化的过程就是用这种方式开始，并导致越来越复杂的自我复制组织的产生。第一种原始的生命形式消化了包括硫化氢在内的不同物质，而释放出氧气。这就逐渐地将大气改变成今天这样的成分，并且允许诸如鱼、爬行动物、哺乳动物以及最后人类等生命的更高形式的发展。

宇宙从非常热的状态开始并随膨胀而冷却的景象，和我们今天所有的观测证据相一致。尽管如此，它还留下许多未被回答的重要问题：

（1）为何早期宇宙如此之热？

（2）为何宇宙在大尺度上如此均匀？为何它在空间的所有点上和所有方向上看起来相同？尤其是，当我们朝不同方向看时，为何微波辐射背景的温度几乎完全相同？这有点像问许多学生一个考试题。如果所有人都给出完全相同的回答，你就会相当肯定，他们相互之间交流过。在上述的模型中，从大爆炸开始光还没有来得及从一个遥远的区域到达另一个区域，即使这两个区域在宇宙的早期靠得很近。按照相对论，如果连光都不能从一个区域到达另一个区域，则没有任何其他的信息能做到。所以，除非因为某种不能解释的原因，导致早期宇宙中不同的区域刚好从同样的温度开始，否则没有一种方法能使它

们达到相互一样的温度。

（3）为何宇宙以这么接近于区分坍缩和永远膨胀模型的临界膨胀率开始，这样即使在100亿年以后的现在，它仍然几乎以临界的速率膨胀？如果在大爆炸后的1秒钟那一时刻其膨胀率哪怕小十亿亿分之一，那么在它达到今天这么大的尺度之前宇宙早已坍缩。

（4）尽管宇宙在大尺度上是如此的一致和均匀，它却包含有局部的无规性，诸如恒星和星系。人们认为，这些是从早期宇宙中不同区域之间密度的细小差别发展而来的。这些密度起伏的起源是什么？

广义相对论本身不能解释这些特征或回答这些问题，因为它预言，宇宙是从在大爆炸奇点处的无限密度起始的。广义相对论和所有其他物理定律在奇点处都失效了：人们不能预言从奇点会出来什么。正如以前解释的，这表明我们可以从这理论中割除去大爆炸奇点和任何先于它的事件，因为它们对我们没有任何观测效应。时空会有一个边界——大爆炸处的开端。

科学似乎揭示了一组定律，在不确定性原理设下的极限内，如果我们知道宇宙在任一时刻的状态，这些定律就会告诉我们，它如何随时间发展。这些定律也许原先是由上帝颁布的，但是看来从那以后他就让宇宙自身按照这些定律去演化，而现在不对它干涉。但是，他是怎么选择宇宙的初始状态和结构的呢？什么是在时间起始处的“边界条件”？

一种可能的回答是，上帝选择宇宙的这种初始结构是因为某些我们无望理解的原因。这肯定是在一个全能造物主的力量之内。但是如果他使宇宙以这种不能理解的方式开始，他为何又选择让它按照我们可理解的定律去演化？整部科学史正是对事件不是以任意方式发生，而是反映了一定内在秩序的逐步的意识。这秩序可以是，也可以不是由神灵启示的。只有假定这种秩序不但应用于定律，而且应用于时空

边界处的条件时才是自然的，这种条件指明宇宙的初始态。可以有大量具有不同初始条件的宇宙模型，它们都服从定律。应该存在某种原则去抽取一个初始状态，也就是一个模型，去代表我们的宇宙。

所谓的混沌边界条件即是这样一种可能性。这些条件含蓄地假定，要么宇宙是空间无限的，要么存在无限多宇宙。在混沌边界条件下，在刚刚大爆炸之后，寻求任何空间区域在任意给定的结构的概率，在某种意义上，和它在任何其他结构的概率是一样的：宇宙初始态的选择纯粹是随机的。这意味着，早期宇宙可能是非常混沌和无序的。因为与光滑和有序的宇宙相比，存在着多得多的混沌和无序的宇宙。（如果每一结构都是等几率的，因为混沌无序态多得这么多，宇宙多半会从这种态起始）。很难理解，从这样混沌的初始条件，如何导致今天我们这个在大尺度上如此光滑和规则的宇宙。人们还预料，在这样的模型中，密度起伏导致比伽马射线背景观测设定的上限多得多的太初黑洞的形成。

如果宇宙确实是空间无限的，或者如果存在无限多宇宙，就会存在某些从光滑和一致的形态开始演化的大的区域。这有点像著名的一大群猴子锤击打字机的故事——它们所写的大部分都是废话。但是纯粹由于偶然，它们可能碰巧打出莎士比亚的一首十四行诗。类似地，在宇宙的情形下，是否我们可能刚好生活在一个光滑和均匀的区域里呢？初看起来，这是非常不可能的，因为这样光滑的区域比混沌的无序的区域稀罕得多。然而，假定只有在光滑的区域里星系、恒星才能形成，才能有合适的条件，让像我们这样复杂的，能自然复制的机体得以发展，这种机体能够质疑宇宙为什么如此光滑的问题。这就是应用称为人存原理的一个例子。人存原理可以解释为：“我们看到的宇宙之所以如此，乃是因为我们的存在。”

人存原理有弱的和强的意义下的两种版本。弱人存原理是讲，在一个大的或具有无限空间和/或时间的宇宙里，只有在某些时空有限的区域里，才存在智慧生命发展的必要条件。因此，在这些区域中，

如果智慧生物观察到他们在宇宙的位置满足他们存在必要的条件，他们就不应感到惊讶。这有点像生活在富裕街坊的富人看不到任何贫穷。

应用弱人存原理的一个例子是“解释”为何大爆炸发生于大约100亿年之前——智慧生物大约需要那么长时间演化。正如前面解释的，一个早代的恒星必须首先形成。这些恒星将原先的一些氢和氦转化成像碳和氧这样的元素，由这些元素构成我们。然后恒星作为超新星而爆发，其裂片形成其他恒星和行星，其中就包括我们的太阳系，太阳系年龄大约是50亿年。地球存在的头10亿或20亿年，对于任何复杂东西的发展都嫌太热。余下的30亿年左右才用于生物进化的漫长过程，从最简单的生命，直到能够测量回溯到大爆炸的时间的生命，就在此期间形成。

很少有人会对弱人存原理的有效性提出异议。然而，有的人走得更远并提出强人存原理。按照这个理论，要么存在许多不同的宇宙，要么存在一个单独宇宙的许多不同的区域，每一个都有自己初始的结构，或许还有自己的一组科学定律。这些宇宙中的大多数，不具备复杂机体发展的合适条件；只有在少数像我们的宇宙中，智慧生命才得以发展并能质疑：“为何宇宙是我们看到的这种样子？”答案很简单：如果它不是这个样子，我们就不会在这里！

我们现在知道，科学定律包含许多基本的数，如电子电荷的大小以及质子和电子的质量比。至少现在，我们不能从理论上预言这些数值——我们必须由观测找到它们。也许有一天，我们会发现一个将它们所有都预言出来的完备的统一理论，但是还有可能它们之中的一些或全部，在不同的宇宙或在一个单一宇宙之中是变化的。值得注意的事实是，这些数值看来是被非常细微地调整到让生命得以发展。例如，如果电子的电荷只要稍微有点不同，则要么恒星不能够燃烧氢和氦，要么它们没有爆炸过。当然，也许存在其他形式的、甚至还没被科学幻想作家梦想过的智慧生命。它并不需要像太阳这样恒星的光，

或在恒星中制造出并在它爆炸时被抛到空间去的更重的化学元素。尽管如此，看来很清楚，允许任何智慧生命形式的发展的数值范围是比较小的。对于大部分数值的集合，宇宙也会产生，虽然它们可以是非常美的，可惜不包含任何一个能为如此美丽而倾倒的人。人们既可以认为这是在创生和科学定律选取中的神意的证据，也可以认为是对强人存原理的支持。

人们可以提出一系列理由，来反对用强人存原理解释观察到的宇宙状态。首先，在何种意义上，可以说所有这些不同的宇宙存在？如果它们确实相互隔开，在其他宇宙中发生的事件在我们自己的宇宙中就没有可观测的后果。所以，我们应该用经济原理，将它们从理论中割除掉。另一方面，它们若仅仅是一个单一宇宙的不同区域，则在每个区域里的科学定律必须是一样的，否则人们就不能从一个区域连续地运动到另一区域。在这种情况下，不同区域之间的仅有的不同是它们的初始结构。这样，强人存原理即归结为弱人存原理。

对强人存原理的第二个异议是，它和整个科学史的潮流背道而驰。我们现代的图象是从托勒密和他的支持者的地心宇宙论出发，通过哥白尼和伽利略日心宇宙论发展而来的。在此图象中，地球是一个中等大小的行星，它围绕着一个寻常的螺旋星系外圈的普通恒星作公转，而这星系本身只是可观察到的宇宙中大约万亿个星系之一。然而强人存原理却宣布，这整个庞大的构造仅仅是因我们的缘故而存在，这是非常令人难以置信的。我们太阳系肯定是我们存在的前提，人们可以将之推广于我们的整个星系，作为让产生重元素的早代恒星存在的前提。但是，丝毫看不出存在任何其他星系的必要，宇宙在大尺度上也不必在每一方向上必须如此一致和类似。

如果人们能够表明，宇宙的相当多不同的初始结构会演化产生像我们今天看到的宇宙，至少在弱的形式上，人们会对人存原理感到更满意。如果果真如此，则一个从某些随机的初始条件发展而来的宇宙，应当包含许多光滑均匀的区域，而且这些区域适合智慧生命演

化。另一方面，如果必须极端仔细地选择宇宙的初始条件，才能导致在我们周围所看到的一切，宇宙就不太可能包含任何会出现生命的区域。在上述的热大爆炸模型中，热来不及从一个区域流到另一区域。这意味着宇宙的初始态在每一处必须刚好有同样的温度，才能说明我们在每一方向上看到的微波背景辐射都有同样温度。其初始的膨胀率也要非常精确地选择，才能使现在的膨胀率仍然这么接近于需要用以避免坍缩的临界速率。这表明，如果热大爆炸模型直到时间的开端都是正确的，则确实必须非常仔细地选择宇宙的初始态。所以，除非作为上帝有意创造像我们这样生命的行为，否则很难解释，为何宇宙只用这种方式起始。

为了试图寻找一个能从许多不同的初始结构演化到像现在这样的宇宙的东西，麻省理工学院的科学家阿伦·固斯提出，早期宇宙可能经历过一个非常快速膨胀的时期。这种膨胀叫做“暴胀”，意指宇宙在一段时间里，不像现在这样以减少的，而是以增加的速率膨胀。按照固斯理论，在远远小于1秒的时间里，宇宙的半径增大了100万亿亿亿（1后面跟30个0）倍。

固斯提出，宇宙是以一种非常热而且相当混沌的状态从大爆炸起始的。这些高温表明宇宙中的粒子运动得非常快并具有高能量。正如早先我们讨论过的，人们预料在这么高的温度下，强和弱核力及电磁力都被统一成一个单独的力。随着宇宙膨胀，它会变冷，而粒子能量下降。最后出现了所谓的相变，并且力之间的对称性被破坏了：强力变得和弱力以及电磁力不同。相变的一个普通的例子是，当水降温时会冻结成冰。液态水是对称的，它在任何一点和任何方向上都是相同的。然而，当冰晶体形成时，它们有确定的位置，并在某一方向上整齐排列。这就破坏了水的对称。

在水的情形，只要你足够小心，就能使之“过冷”：也就是可以将温度降低到冰点（0℃）以下而不结冰。固斯认为，宇宙的行为也很相似：宇宙温度可以降低到临界值以下，而各种力之间的对称没有受

到破坏。如果发生这种情形，宇宙就处于一个不稳定状态，其能量比对称破缺时更大。可以指出，这特殊的额外能量呈现出反引力的效应：其作用如同一个宇宙常数。宇宙常数是当爱因斯坦在试图建立一个稳定的宇宙模型时，引进广义相对论之中去的。由于宇宙已经像大爆炸模型那样膨胀，所以这宇宙常数的排斥效应使得宇宙以不断增加的速度膨胀。即使在一些物质粒子比平均数更多的区域，这一有效宇宙常数的排斥作用也超过了物质的引力吸引作用。这样，这些区域也以加速暴胀的形式膨胀。当它们膨胀时，物质粒子就越分越开，留下了一个几乎不包含任何粒子，并仍然处于过冷状态的膨胀的宇宙。这种膨胀抹平了宇宙中的任何不规则性，正如当你吹胀气球时，它上面的皱纹就被抹平了。这样，从许多不同的非均匀的初始状态可以演化出宇宙现在光滑均匀的状态。

在这样一个其膨胀由宇宙常数加速，而不因物质的引力吸引使之减慢的宇宙中，早期宇宙中的光线就有足够的时间从一个区域旅行到另一个区域。这就解答了早先提出的，为何在早期宇宙中的不同区域具有同样性质的问题。不但如此，宇宙的膨胀率也自动变得非常接近由宇宙的能量密度决定的临界值。这就能够解释，不需假设宇宙初始膨胀率曾被非常仔细地选择过，为何现在的膨胀率仍然这么接近临界值。

暴胀的思想还能解释为何在宇宙中存在这么多物质。在我们能观察到的宇宙中大约有1亿亿亿亿亿亿亿亿亿（1后面跟80个0）个粒子。它们从何而来？答案是，在量子理论中，粒子了以以粒子/反粒子对的形式由能量中创生出来。但这只不过引起能量从何而来的问题。答案是，宇宙的总能量准确为零。宇宙中的物质是由正能量产生的。然而，物质本身由于引力总是吸引的。两块相互靠近的物质比两块分得很开的物质具有较少的能量，因为你必须消耗能量去克服把它们拉在一起的引力才能将其分开。这样，在一定意义上，引力场具有负能量。在空间上大体一致的宇宙的情形中，人们可以证明，这个负

的引力能刚好抵消了物质所代表的正能量。这样，宇宙的总能量为零。

零的2倍仍为零。这样，宇宙可以同时将其正的物质能和负的引力能加倍，而不违反能量守恒。在宇宙正常膨胀时，这并没有发生。这时当宇宙变大时，物质能量密度下降。然而，这种情形确实发生于暴胀时期。因为当宇宙膨胀时，过冷态的能量密度保持不变：当宇宙体积加倍时，正物质能和负引力能都加倍，这样总能量保持为零。在暴胀相，宇宙的尺度增大了一个非常大的倍数。这样，可用以制造粒子的总能量变得非常大。正如固斯说过的："都说没有免费午餐这回事，但是宇宙却是最彻底的免费午餐。"

今天宇宙不是以暴胀的方式膨胀。这样，必须有一种机制，它可以消去这一非常大的有效宇宙常数，从而使膨胀率从加速的状态改变为如同今天这样由引力减慢的状态。人们可以预料，在宇宙暴胀时各种力之间的对称最终会破缺，正如过冷的水最终会凝固一样。这样，未破缺的对称态的额外能量就会释放，并将宇宙重新加热到刚好低于使各种力对称的临界温度。以后，宇宙就以标准的大爆炸模式继续膨胀并变冷。但是，现在我们可以解释，为何宇宙刚好以临界速率膨胀，并且为何不同的区域具有相同的温度。

在固斯的原先设想中，有点像在非常冷的水中出现冰晶体，相变是突然发生的。其想法是，正如同沸腾的水围绕着蒸汽泡，新的对称破缺相的"泡泡"在原有的对称相中形成。设想泡泡膨胀并相互碰撞，直到整个宇宙处于新相。麻烦在于，正如同我和其他几个人指出的，宇宙膨胀得如此之快，即使泡泡以光速胀大，它们也要相互分离，并因此不能合并在一起。结果宇宙变成一种非常不均匀的状态，有些区域仍具有各种力之间的对称。这样的模型跟我们观察到的宇宙不吻合。

1981年10月，我去莫斯科参加量子引力的会议。会后，我在斯

特堡天文研究所做了一个有关暴胀模型和它的问题的讲演。在此之前，我请其他人替我宣读讲稿，因为大多数人听不懂我的声音。但是这一次我来不及准备讲稿，所以我自己讲，让我的一名研究生逐字逐句地重复我的话。演讲进行得很顺利，并且使我有多得多的时间和听众交谈。听众席中有一位年轻的苏联人，莫斯科列别捷夫研究所的安德雷·林德。他说，如果泡泡是如此之大，使得我们的宇宙区域被整个地包含在一个单独的泡泡之中，则可以避免泡泡不能合并在一起的困难。为了使这个行得通，从对称相向对称破缺相的改变必须在泡泡中发生得非常缓慢，但是按照大统一理论这是完全可能的。林德的缓慢对称破缺思想非常好，但是过后我意识到，他的泡泡在那一时刻必须比宇宙的尺度还要大！我指出，那时对称不仅仅在泡泡里，而且在所有的地方同时被破坏。这会导致一个正如我们观察到的一致的宇宙。我被这个思想弄得非常激动，并和我的一个学生因·莫斯讨论。然而，当我后来收到一个科学杂志社寄来的林德的论文，征求是否可以发表时，作为他的朋友，我感到相当难为情。我答复说，这里有一个关于泡泡比宇宙还大的瑕疵，但是里面关于缓慢对称破缺的基本思想是非常好的。我建议将此论文照原样发表。因为林德要花几个月时间去改正它，并且他寄到西方的任何东西都要通过苏联的审查，这种对于科学论文的审查既无技巧可言又很缓慢。我和因·莫斯便越俎代庖，为同一杂志写了一篇短文。我们在该文中指出这泡泡的问题，并提出如何将其解决。

我从莫斯科返回的第二天，即去费城接受富兰克林研究所的奖章。我的秘书朱迪·费拉施展其不俗的魅力说服了英国航空公司给她自己和我免费提供协和式飞机的广告旅行坐席。然而，在去机场的路上被大雨耽搁，我没赶上航班。尽管如此，我最终还是到了费城并得到奖章。之后，我应邀在费城的爵索尔大学作了关于暴胀宇宙的演讲。我所作的讲演，正和在莫斯科的一样，是关于暴胀宇宙的问题。

几个月之后，宾州大学的保罗·斯特恩哈特和安德鲁斯·阿伯勒希

特独立地提出和林德非常相似的思想。现在他们和林德分享以缓慢对称破缺的思想为基础的所谓“新暴胀模型”的荣誉。（旧的暴胀模型是指固斯关于形成泡泡后快速对称破缺的原始设想。）

新暴胀模型是一个好的尝试，它能解释宇宙为何是这种样子。然而我和其他几个人指出，至少在它原先的形式，它预言的微波背景辐射的温度变化要比观察到的大得多。后来的工作还对极早期宇宙中是否存在过这类需要的相变提出怀疑。我个人的意见是，现在新暴胀模型作为一个科学理论气数已尽。虽然还有很多人似乎不承认它的死亡，还继续写文章，好像那理论还有生命力。1983年，林德提出了一个更好的所谓混沌暴胀模型。这里没有相变和过冷，而代之以存在一个自旋为0的场，由于它的量子涨落，在早期宇宙的某些区域有大的场值。在那些区域中，场的能量起到宇宙常数的作用，它具有排斥的引力效应，而使这些区域以暴胀的形式膨胀。随着它们膨胀，它们中的场的能量慢慢地减小，直到暴胀改变到犹如热大爆炸模型中的膨胀时为止。这些区域之一就成为可观察的宇宙让我们看到。这个模型具有早先暴胀模型的所有优点，但是它并不取决于使人生疑的相变，此外，它还能给出微波背景辐射温度起伏的合理幅度，这与观测相符合。

暴胀模型的这个研究指出：宇宙现在的状态可以从相当大量的不同初始结构引起。这很重要，因为它表明不必非常细心地选取我们居住的那部分宇宙区域的初始状态。所以，如果愿意的话，我们可以利用弱人存原理解释宇宙为何现在如此这般。然而，绝不是任何一种初始结构都会产生像我们观察到的宇宙。这一点很容易做到。考虑现在宇宙处于一个非常不同的态，例如一个非常成团的非常无规则的态。人们可以利用科学定律，在时间上将其演化回去，以确定宇宙在更早时刻的结构。按照经典广义相对论的奇点定理，仍然存在一个大爆炸奇点。如果你在时间前进方向上按照科学定律演化这样的宇宙，你就会得到你从其开始的那个成团的无规则的态。这样，必定存在不会产

生像我们今天观察到的宇宙的初始结构。所以，就连暴胀模型也没有告诉我们，为何初始结构不是那种态，从它演化成和我们观测到的非常不同的宇宙。我们是否应该再从人存原理得到解释呢？难道所有这一切仅仅是因为好运气？看来，这只是无望的遁词，是对我们理解宇宙根本秩序的所有希望的否定。

为了预言宇宙应该如何起始，人们需要在时间开端处成立的定律。罗杰·彭罗斯和我证明的奇点定理指出，如果广义相对论的经典理论是正确的，则时间的开端是具有无限密度和无限时空曲率的一点，在这样的点上所有已知的科学定律都崩溃。人们可以设想存在在奇点处成立的新定律，但是在如此不守规矩之处，甚至连表述这样的定律都是非常困难的，而且从观察中我们没有得到关于这些定律应是什么样子的任何指示。然而，奇点定理真正揭示的是，引力场变得如此之强，使量子引力效应变得十分重要：经典理论已经不能很好地描述宇宙。这样，人们必须用量子引力论去讨论宇宙的极早期阶段。正如我们将会看到的，在量子力学中，通常的科学定律有可能在任何地方都有效，包括时间开端这一点在内：不必针对奇点提出新的定律，因为在量子理论中不必存在任何奇点。

我们仍然没有一套完备而协调的理论将量子力学和引力结合在一起。然而，我们相当清楚这样一套统一理论所应该具备的某些特征。其中一个就是它必须和费恩曼提出的按照对历史求和的量子力学表述相合并。在这种方法里，一个粒子不像在经典理论中那样，不仅只有一个单独的历史。相反，它被认为通过时空里的任何可能的路径，这些历史中的每一个都有一对相关的数，一个代表波的幅度，另一个代表它在循环中的位置（相位）。粒子通过某一特定点的概率是将通过此点的所有可能历史的波叠加求得。然而，当人们实际去进行这些求和时，就遭遇到了严重的技术问题。回避这个问题的仅有的独特方法是：你必须不是对发生在你我经验的“实的”时间内的，而是对发生在所谓“虚的”时间内的粒子历史的波进行求和。虚时间可能听起来像是

科学幻想，但事实上，它是定义得很好的数学概念。如果你取任何平常的（或“实的”）数和它自己相乘，结果是一个正数（例如2乘2是4，但-2乘-2也是这么多。）然而，存在一种特别的数（叫虚数），当它们自乘时得到负数（叫做i的数自乘时得-1，2i自乘得-4，等等）。

人们可以用下面的办法来图解实数和虚数：实数可以用一根从左至右的线来代表，中间是零点，像-1，-2等负数在左边，而像1，2等正数在右边。而虚数由书页上一根上下的线来代表，i，2i等在中点以上，而-i，-2i等在中点以下。这样，在某种意义上，虚数和通常的实数夹一直角。

人们必须利用虚时间，以避免在进行费恩曼对历史求和的技术上的困难。也就是说，为了计算的目的，人们必须用虚数而不是用实数来测量时间。这对时空有一有趣的效应：时间和空间的区别完全消失。事件具有虚值时间坐标的时空称为欧几里得型的，它是采用建立了二维面几何的希腊人欧几里得的名字命名的。我们现在称之为欧几里得时空的东西，除了是四维而不是二维以外，其余的和它都非常相似。在欧几里得时空中，时间方向和在空间中的方向没有不同之处。另一方面，在通常用实的时间坐标来标记事件的实的时空里，人们很容易区别这两种方向——位于光锥中的任何点是时间方向，位于光锥之外的为空间方向。无论如何，就日常的量子力学而言，我们利用虚的时间和欧几里得时空，可以认为仅仅是一个计算有关实时空的答案的数学手段（或技巧）。

我们相信，作为任何终极理论的一部分而不可或缺的第二个特征是爱因斯坦的思想，即引力场由弯曲的时空来代表：粒子在弯曲空间中试图沿着最接近于直线的某种路径走。但是因为时空不是平坦的，它们的路径看起来似乎被引力场折弯了。当我们利用费恩曼的历史求和方法去处理爱因斯坦的引力观点时，和粒子的历史相类似的东西则是代表整个宇宙历史的完整的弯曲时空。为了避免实际进行历史求和的技术困难，这些弯曲的时空必须采用欧几里得型的。也就是，时间

是虚的并和空间的方向不可区分。为了计算找到具有一定性质的，例如在每一点和每一方向上看起来都一样的实时空的概率，人们把和所有具有这性质的历史相关联的波叠加起来即可。

在广义相对论的经典理论中，可能有许多不同的弯曲时空，每一个对应于宇宙不同的初始态。如果我们知道我们宇宙的初始态，我们就会知道它的整个历史。类似地，在量子引力论中，宇宙可能存在许多不同的量子态。同样地，如果我们知道在历史求和中的欧几里得弯曲时空在早先时刻的行为，我们就会知道宇宙的量子态。

在以实的时空为基础的经典引力论中，宇宙可能的行为只有两种方式：要么它已存在了无限长时间，要么它在有限的过去的某一时刻的奇点上有一个开端。而在量子引力论中，产生了第三种可能性。因为人们采用欧几里得时空，在这里时间方向和空间方向具有相同的地位，所以时空有可能在范围上是有限的，却没有形成边界或边缘的奇点。时空就像是地球的表面，只不过多了两维。地球的表面在范围上是有限的，但它没有边界或边缘：如果你朝着落日的方向驾船，你不会掉到边缘外面或陷入奇点中去。（因为我曾经环球旅行过，所以知道！）

如果欧几里得时空延伸到无限的虚时间，或者在一个虚时间奇点处开始，我们就会遇到和经典理论中指定宇宙初态同样的问题：即上帝可以知道宇宙如何开始，但是我们提不出任何特别原因，认为它应以这种而不是那种方式开始。另一方面，量子引力论开辟了另一种新的可能性，在这里时空没有边界，所以没有必要指定边界上的行为。这里不存在在该处科学定律崩溃的奇点，也就是不存在在该处必须祈求上帝或某些新的定律给时空设定边界条件的时空边缘。人们可以说：“宇宙的边界条件是它没有边界。”宇宙便是完全自足的，而不受任何外在于它的东西影响。它既不被创生，也不被消灭。它就是存在。

我正是在早先提到的那次梵蒂冈会议上首次提出，时间和空间可能会共同形成一个在尺度上有限却没有任何边界或边缘的面。然而我的论文数学气息太浓，所以文章中包含的上帝在创生宇宙的作用的含义在当时没被普遍意识到（对我也是如此）。在梵蒂冈会议期间，我不知道如何用“无边界”思想去预言宇宙。然而，次年夏天我在加州大学的圣他巴巴拉分校度过。在那里，我的一位朋友兼合作者詹姆·哈特尔和我共同得出了如果时空没有边界时宇宙应满足的条件。回到剑桥后，我和我的两个研究生朱丽安·拉却尔和约纳逊·哈里威尔继续从事这项工作。

我要着重说明，时空是有限而“无界”的思想仅仅是一个设想，它不能从其他原理导出。正如任何其他科学理论，它原先可由美学或形而上学的原因提出，但是它给出的预言是否与观测相一致是对它的真正检验。不过，在量子引力的情况下，由于以下两个原因这很难确定。首先，正如将要在第十一章解释的，虽然我们对能将广义相对论和量子力学合并在一起的理论应具有的方式，已经知道得相当多，但是还不能准确地知道哪种理论能成功地做到这一点。其次，任何详尽描述整个宇宙的模型在数学上都过于复杂，使我们不能通过计算做出准确的预言。所以，人们不得不做简化的假设和近似——并且甚至这样，要从中引出预言仍是令人生畏的课题。

对历史求和中的每一个历史不只描述时空，而且描述在其中的任何东西——包括像能观察宇宙历史的人类那样复杂的生物。这可对人存原理提供另一个支持，因为如果任何历史都是可能的，就可以用人存原理去解释为何我们发现宇宙是当前这样子。对我们并不存在其中的其他历史究竟应赋予什么意义还不清楚。然而，如果利用对历史求和可以显示，我们的宇宙不只是一个可能的，而且是最有可能的历史，则这个量子引力论的观点就会令人满意得多。为此，我们必须对所有可能的没有边界的欧几里得时空进行历史求和。

人们从“无边界”假定得知，宇宙遵循大多数历史的机会是可以忽

略不计的，但是有一组特别的历史比其他的历史有多得多的机会。这些历史可以描绘得像地球的表面。在那里与北极的距离代表虚的时间，并且离北极等距离的圆周长代表宇宙的空间尺度。宇宙作为单独一点从北极起始。随着人们往南走，离开北极等距离的纬度圈变大，这和宇宙随虚时间的膨胀相对应（图8.1）。宇宙在赤道处会达到最大的尺度，并且随着虚时间的继续增加而收缩，最后在南极收缩成一点。尽管宇宙在南北二极的尺度为零，但是这些点不是奇点，它们并不比地球上的北南二极更奇异。科学定律在它们那里有效，正如同它们在地球上的南北二极有效一样。

然而，宇宙的历史在实的时间里显得非常不一样。大约在100亿年或200亿年以前，它有一个最小的尺度，它等于历史在虚时间里的最大半径。在后来的实时间里，宇宙就会像由林德设想的混沌暴胀模型那样地膨胀（但是现在人们不必假定宇宙以某种方式从一类合适的状态产生出来）。宇宙会膨胀到一个非常大的尺度，并最终重新坍缩成为在实时间里看起来像是奇点的一个东西。这样，在某种意义上说，即使我们躲开黑洞，仍然是注定要毁灭的。只有当我们按照虚时间来描绘宇宙时才不会有奇点。

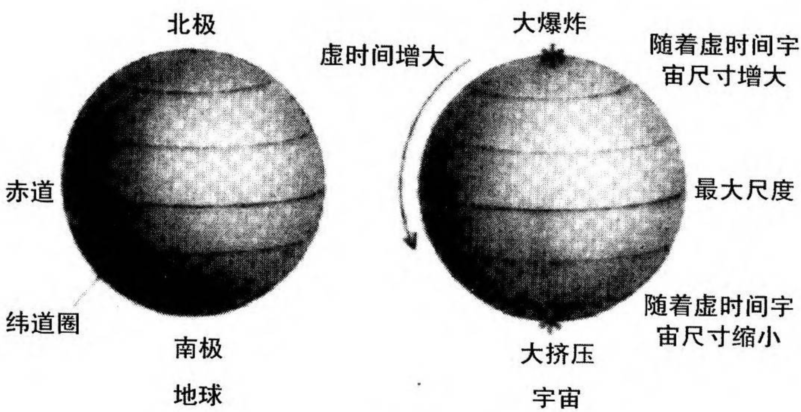



图8.1

如果宇宙确实处在这样的一个量子态，宇宙在虚时间里就没有奇点。因此，我近期的工作似乎使我早期研究奇点的工作成果完全付之东流。但是正如上面指出的，奇点定理的真正重要性在于，它们指出引力场必然会强到不能无视量子引力效应的程度。这接着导致也许在虚时间里宇宙的尺度有限但没有边界或奇点的观念。然而，当人们回到我们生活其中的实时间时，那里仍会出现奇点。陷进黑洞的那位可怜的航天员的结局仍然是极可悲的；只有当他在虚时间里生活，才不会遭遇到奇点。

上述这些也许暗示所谓的虚时间才是真正的实时间，而我们叫做实时间的东西恰恰是子虚乌有的空想的产物。在实时间中，宇宙具有开端和终结的奇点，这奇点构成了科学定律在那里失效的时空边界。但是，在虚时间里不存在奇点或边界。所以，很可能我们称作虚时间的才真正是更基本的观念，而我们称作实时间的反而是我们臆造的，它仅仅有助于我们描述我们认为的宇宙模样，如此而已。但是，按照我在第一章描述的方法，科学理论只不过是我们用以描述自己观察的数学模型：它只存在于我们的头脑中。所以去问诸如这样的问题是毫无意义的："实的"或"虚的"时间，哪一种是实在的？这仅仅是哪一种描述更为有用的问题。

人们还可以利用对历史求和以及无边界设想去发现宇宙的哪些性质很可能发生。例如，人们可以计算，当宇宙具有现在密度的某一时刻，在所有方向上以几乎同等速率膨胀的概率。在迄今已被考察的简化的模型中，发现这个概率是高的；也就是说，无边界设想导致一个预言，即宇宙现在在每一方向的膨胀率几乎相同是极其可能的。这与微波背景辐射的观测相一致，它指出在任何方向上具有几乎完全同样的强度。如果宇宙在某些方向比其他方向膨胀得更快，一个附加的红移就会减小那些方向辐射的强度。

人们正在研究无边界条件的进一步预言。一个特别有趣的问题是，早期宇宙中物质密度对其均匀密度的小幅度偏离。这些偏离首先

引起星系，然后是恒星，最后是我们自身的形成。不确定性原理意味着，早期宇宙不可能是完全均匀的，因为粒子的位置和速度必定存在一些不确定性或起伏。利用无边界条件，我们发现，在事实上，宇宙必须恰好从由不确定性原理允许的最小可能的非均匀性开始。然后，正如在暴胀模型中预言的一样，宇宙经历了一段快速膨胀时期。在这个期间，初始的非均匀性被放大到足以解释在我们周围观察到的结构的起源。1992年宇宙背景探险者卫星（COBE）首次检测到微波背景随方向的非常微小的变化。这种非均匀性随方向的变化方式似乎和暴胀模型以及无边界设想的预言相符合。这样，在卡尔，波普的意义上，无边界设想是一种好的科学理论：它可以被观测证伪，但是它的预言却被证实了。在一个各处物质密度稍有变化的膨胀宇宙中，引力使得较紧密区域的膨胀减慢，并使之开始收缩。这就导致星系、恒星和最终甚至像我们自己这样微不足道的生物的形成。这样，宇宙无边界条件和量子力学中的不确定性原理一道，可以解释我们在宇宙中看到的所有复杂的结构。

空间和时间可以形成一个没有边界的闭曲面的思想，对于上帝在宇宙事务中的作用还有一个深远的含义。随着科学理论在描述事件方面的成功，大部分人进而相信上帝允许宇宙按照一套定律来演化，而不介入其间使宇宙触犯这些定律。然而，定律并没有告诉我们，宇宙的太初应该像什么样子——它依然要靠上帝去卷紧发条，并选择如何去启动它。只要宇宙有一个开端，我们就可以设想存在一个造物主。但是，如果宇宙的的确确是完全自足的，没有边界或边缘，它就既没有开端也没有终结：它就是存在。那么，还会有造物主存身之处吗？

# 第九章 时间箭头

我们在前几章中看到了，长期以来人们关于时间性质的观点是如何变化的。直到本世纪初（即20世纪初——编者注），人们还相信绝对时间。也就是说，每一事件可由一个称为“时间”的数以惟一的方式来标记，所有好的钟在测量两个事件之间的时间间隔上都是一致的。然而，对于任何正在运动的观察者，不管他怎么运动，光速总是一样的这一发现，导致了相对论——而在相对论中，人们必须抛弃存在一个惟一的绝对时间的观念，相反，每个观察者都有由他携带的钟记录的他自己的时间测量：不同观察者携带的钟不必要一致。这样，相对于进行测量的观察者而言，时间变成一个更个人的概念。

当人们试图统一引力和量子力学时，必须引入“虚”时间的概念。虚时间是不能和空间方向区分的。如果一个人能往北走，他就能转过头并朝南走；同样的，如果一个人能在虚时间里向前走，他应该能够转过来并往后走。这表明在虚时间里，往前和往后之间不可能有重要的差别。另一方面，当人们考察“实”时间时，正如众所周知的，在前进和后退方向存在着非常巨大的差别。过去和将来之间的这种差别从何而来？为何我们记住过去而不是将来？

科学定律并不区别过去和将来。更精确地讲，正如前面解释的，科学定律在称作C、P和T的联合作用（或对称）下不变。（C是指用反粒子替代粒子。P的意思是取镜像，这样左和右就相互交换了。而T是指颠倒所有粒子的运动方向：事实上，是使运动倒退回去。）在所有正常情形下，制约物体行为的科学定律在CP联合对称下独自不变。

换言之，对于其他行星上的居民，若他们是我们的镜像并且由反物质而不是物质构成，则生活会刚好和我们一样。

如果科学定律在CP联合作用以及CPT联合作用下都不变，它们也必须在单独的T作用下不变。然而，在日常生活的实时间中，前进和后退的方向之间还是有一个大的差异。想像一杯水从桌子上滑落下，在地板上被打碎。如果你将其录像，你可以容易地辨别出它是向前进还是向后退。如果将其倒放回来，你会看到碎片忽然集中到一起离开地板，并跳回到桌子上形成一个完整的杯子。你可断定录像是在倒放，因为在日常生活中从未见过这种行为。如果发生这样的事，陶瓷业将无生意可做。

为何我们从未看到破碎的杯子集合起来，离开地面并跳回到桌子上，通常的解释是这违背了热力学第二定律。它可表述为，在任何闭合系统中无序度或熵总是随时间而增加。换言之，它是穆菲定律的一种形式：事情总是越变越糟！桌面上一个完整的杯子是一个高度有序的状态，而地板上破碎的杯子是一个无序的状态。人们很容易从早先桌子上的杯子变成后来地面上的碎杯子，而不是相反。

无序度或熵随着时间增加是所谓的时间箭头的一个例子。时间箭头将过去和将来区别开来，使时间有了方向。至少有三种不同的时间箭头：第一个，是热力学时间箭头，即是在这个时间方向上无序度或熵增加；然后是心理学时间箭头，这就是我们感觉时间流逝的方向，在这个方向上我们可以记忆过去而不是未来；最后，是宇宙学时间箭头，宇宙在这个方向上膨胀，而不是收缩。

我将在这一章论断，宇宙的无边界条件和弱人存原理一起能解释为何所有的三个箭头指向同一方向——此外，为何必须存在一个定义得很好的时间箭头。我将论证热力学箭头确定心理学箭头，并且这两种箭头必须总是指向相同的方向。如果人们假定宇宙的无边界条件，我们将看到必然存在定义得很好的热力学和宇宙学时间箭头。但对于

宇宙的整个历史来说，它们并不总是指向同一方向。然而，我将论断，只有当它们指向一致时，对于能够发问为何无序度在宇宙膨胀的时间方向上增加的智慧生命的发展，才有合适的条件。

首先，我要讨论热力学时间箭头。总存在着比有序状态多得多的无序状态的这一事实，导致热力学第二定律。譬如，考虑一盒拼板玩具，存在一个并且只有一个使这些小纸片拼成一幅完整图画的排列。另一方面，存在巨大数量的排列，这时小纸片是无序的，不能拼成一幅画。

假设一个系统从这少数的有序状态之一出发。随着时间流逝，这个系统将按照科学定律演化，而且它的状态将改变。到后来，因为存在着更多的无序状态，它处于无序状态的可能性比处于有序状态更大。这样，如果一个系统服从一个高度有序的初始条件，它的无序度就会随着时间的增加而增大。

假定拼板玩具盒的纸片从排成一幅图画的有序组合开始，如果你摇动这盒子，这些纸片将会采用其他组合，这很可能是一个不形成一幅合适图画的无序的组合，就是因为存在如此之多得多的无序的组合。有一些纸片团仍可能形成部分图画，但是你越摇动盒子，这些团就越可能被分开，这些纸片将处于完全混乱的状态，在这种状态下它们不能形成任何种类的图画。这样，如果纸片从一个高度有序的状态的初始条件出发，纸片的无序度将可能随时间而增加。

然而，假定上帝决定不管宇宙从何状态开始，它都必须结束于一个高度有序的状态，则在早期这宇宙很可能处于无序的状态。这意味着无序度将随时间而减小。你将会看到破碎的杯子集合起来并跳回到桌子上。然而，任何观察杯子的人都生活在无序度随时间减小的宇宙中，我将论断这样的人会有一个倒溯的心理学时间箭头。这就是说，他们会记住将来的事件，而不是过去的事件。当杯子被打碎时，他们会记住它在桌子上的情形；但是当它在桌子上时，他们不会记住它在

地面上的情景。

由于我们不知道大脑工作的细节，所以谈论人类的记忆是相当困难的。然而，我们确实完全知道计算机的记忆器是如何工作的。因此，我将讨论计算机的心理学时间箭头。我认为，可以合理假定计算机和人类有相同的箭头。如果不是这样，人们可能因为拥有一台记住明天价格的计算机而在股票交易中大发利市！

大体来说，计算机的记忆器是一个包含可处于两种状态中的任一种的元件的设备，算盘是一个简单的例子。其最简单的形式是由许多铁条组成；每一根铁条上有一念珠，此念珠可呆在两个位置中的一个。在计算机记忆器进行存储之前，其记忆器处于无序态，念珠等几率地处于两个可能的状态中（算盘珠杂乱无章地散布在算盘的铁条上。）在记忆器和要记忆的系统相互作用后，根据系统的状态，它肯定处于这种或那种状态（每个算盘珠将要么位于铁条的左边，要么处于右边）。这样，记忆器就从无序态转变成有序态。然而，为了保证记忆器处于正确的状态，需要使用一定的能量（例如，移动算盘珠或给计算机接通电源）。这能量以热的形式耗散了，从而增加了宇宙的无序度的量。人们可以证明，这个无序度增量总比记忆器本身有序度的增量大。这样，由计算机冷却风扇排出的热量表明计算机将一个项目记录在它的记忆器中时，宇宙的无序度的总量仍然增加。计算机记忆过去的时间方向和无序度增加的方向是一致的。

因此，我们对时间方向的主观感觉或心理学时间箭头，是在我们头脑中由热力学时间箭头决定的。正像一台计算机，我们必须在熵增加的顺序上将事物记住。这几乎使热力学定律变成为无聊的东西。无序度随时间的增加乃是因为我们是在无序度增加的方向上测量时间。你不可能有比这个更具胜算的打赌了！

但是究竟热力学时间箭头为何必须存在呢？或换句话说，在我们称之为过去的时间的一端，宇宙为何处于高度有序的状态呢？它为何

不在所有时间里处于完全无序的状态呢？毕竟这似乎更为可能。还有，为何无序度增加的时间方向和宇宙膨胀的方向相同？

在经典广义相对论中，因为所有已知的科学定律在大爆炸奇点处崩溃，人们不能预言宇宙是如何开始的。宇宙可以从一个非常光滑和有序的状态开始。这就会导致正如我们观察到的，定义很好的热力学和宇宙学的时间箭头。但是，它也可以同样合理地从一个非常波浪起伏的无序状态开始。在那种情况下，宇宙已经处于一种完全无序的状态，所以无序度不会随时间增加。它要么保持常数，这时就没有定义得很好的热力学时间箭头；它要么会减小，这时热力学时间箭头就会和宇宙学时间箭头反向。这些可能性的任一种都不符合我们观察到的情况。然而，正如我们看到的，经典广义相对论预言了它自身的崩溃。当时空曲率变大时，量子引力效应变得重要，而经典理论不再能很好地描述宇宙。人们必须用量子引力论去理解宇宙是如何开始的。

正如我们在上一章看到的，在量子引力论中，为了指定宇宙的态，人们仍然必须说清宇宙的可能历史在过去的时空边界是如何行为的。只有当这些历史满足无边界条件，人们才可能避免这个不得不描述我们不知道和无法知道的东西的困难：它们在尺度上有限，但是没有边界、边缘或奇点。在这种情形下，时间的开端就会是规则的光滑的时空的点，并且宇宙在一个非常光滑和有序的状态下开始它的膨胀。它不可能是完全均匀的，否则就违反了量子理论的不确定性原理。必然存在粒子密度和速度的微小起伏，然而无边界条件意味着，这些起伏又是在与不确定性原理相一致的条件下尽可能的小。

宇宙刚开始时有一个指数膨胀或“暴胀”时期，在这期间它的尺度增加了一个非常大的倍数。在这个膨胀过程时，密度起伏起初一直很小，但是后来开始变大。在密度比平均值稍大的区域，额外质量的引力吸引使膨胀速度放慢。最终，这样的区域停止膨胀，并坍缩形成星系、恒星以及像我们这样的生命。宇宙开始时处于一个光滑有序的状态，而随时间演化成波浪起伏的无序的状态。这就解释了热力学时间

箭头的存在。

如果宇宙停止膨胀，并且开始收缩，将会发生什么呢？热力学箭头会不会倒转过来，而无序度开始随时间减少呢？这为从膨胀相存活到收缩相的人们留下了五花八门的类科学幻想的可能性。他们是否会看到杯子的碎片集合起来，离开地板，并跳回到桌子上去呢？他们会不会记住明天的价格，并在股票市场上发财致富？由于宇宙至少要再等100亿年之后才开始收缩，忧虑那时会发生什么似乎有点学究气。但是有一种更快的办法去查明将来会发生什么：跳到黑洞里面去。恒星坍缩形成黑洞的过程和整个宇宙坍缩的后期相当类似。这样，如果在宇宙的收缩相无序度减小，可以预料它在黑洞里面也会减小。所以，一个落到黑洞里去的航天员在下赌注之前，也许能依靠记住轮赌盘上球儿的走向而赢钱。（然而，不幸的是，玩不了多久，他就会变成意大利面条。他也不能使我们知道热力学箭头的颠倒，或者甚至将他赢的钱存入银行，因为他被困在黑洞的事件视界后面。）

起初，我相信在宇宙坍缩时无序度会减小。这是因为，我认为宇宙再次变小时，它必须回到光滑和有序的状态。这表明，收缩相就像是膨胀相的时间反演。处在收缩相的人们将以倒退的方式生活：他们在出生之前即已死去，并且随着宇宙收缩变得更年轻。

这个观念是吸引人的，因为它表明在膨胀相和收缩相之间存在一个漂亮的对称。然而，人们不能置有关宇宙的其他观念于不顾，而只采用这个观念。问题在于：无边界条件是否隐含着这个对称，或者它是否与这个条件不相协调？正如我说过的，我起先以为无边界条件确实意味着无序度会在收缩相中减小。我之所以被误导，部分是由于与地球表面的类比引起的。如果人们将宇宙的开初对应于北极，那么宇宙的终结就应该类似于它的开端，正如南极之与北极相似。然而，南北二极对应于虚时间中的宇宙的开端和终结。在实时间里的开端和终结之间可有非常大的差异。我还被我做过的一项简单的宇宙模型的研究误导，在此模型中坍缩相似乎是膨胀相的时间反演。然而，我的一

位合作者，宾夕法尼亚州立大学的唐·佩奇指出，无边界条件没有要求收缩相必然是膨胀相的时间反演。我的一个学生雷蒙·拉夫勒蒙进一步发现，在一个稍复杂的模型中，宇宙的坍缩和膨胀非常不同。我意识到自己犯了一个错误：无边界条件意味着，事实上在收缩相时无序度继续增加。当宇宙开始收缩时或在黑洞中，热力学和心理学时间箭头不会反向。

当你发现自己犯了像这样的错误后该如何办？有些人从不承认他们是错误的，而继续去找新的往往相互不协调的论据为自己辩解——如爱丁顿在反对黑洞理论时之所为。另外一些人首先宣称，从来没有真正支持过不正确的观点，如果他们支持了，也只是为了显示它是不协调的。在我看来，如果你在出版物中承认自己错了，那会好得多并少造成混乱。爱因斯坦即是一个好的榜样，他在试图建立一个静态的宇宙模型时引入了宇宙常数，他称此为一生中最大的错误。

回头再说时间箭头，余下的问题是：为何我们观察到热力学和宇宙学箭头指向同一方向？或者换言之，为何无序度增加的时间方向正是宇宙膨胀的时间方向？如果人们相信，按照无边界设想似乎隐含的那样，宇宙先膨胀然后重新收缩，那么为何我们应在膨胀相中而不是在收缩相中，这就成为一个问题。

人们可以在弱人存原理的基础上回答这个问题。收缩相的条件不适合智慧人类的存在，而正是他们能够提出这个问题：为何无序度增加的时间方向和宇宙膨胀的时间方向相同？无边界设想预言的宇宙在早期阶段的暴胀意味着，宇宙必须以非常接近为恰好避免坍缩所需要的临界速率膨胀，这样它在很长的时间内才不至于坍缩。到那时候所有的恒星都会烧尽，而在其中的质子和中子可能会衰变成轻粒子和辐射。宇宙将处于几乎完全无序的状态，这时就不会有强的热力学时间箭头。由于宇宙已经处于几乎完全无序的状态，无序度不会增加很多。然而，对于智慧生命的行动来说，一个强的热力学箭头是必需的。为了生存下去，人类必须消耗能量的一种有序形式——食物，并

将其转化成能量的一种无序形式——热量，这样智慧生命不能在宇宙的收缩相中存在。这就解释了为何我们观察到热力学和宇宙学的时间箭头指向一致。并不是宇宙的膨胀导致无序度的增加，而是无边界条件引起无序度的增加，并且只有在膨胀相中才有适合智慧生命的条件。

总之，科学定律并不能区分前进和后退的时间方向。然而，至少存在三个时间箭头，将过去和将来区分开来。它们是热力学箭头，这就是无序度增加的时间方向；心理学箭头，即是在这个时间方向上，我们能记住过去而不是将来；还有宇宙学箭头，也即宇宙膨胀而不是收缩的方向。我指出了心理学箭头本质上应和热力学箭头相同。宇宙的无边界设想预言了存在定义得很好的热力学时间箭头，因为宇宙必须从光滑的有序的状态开始。并且我们看到，热力学箭头和宇宙学箭头的一致，乃是由于智慧生命只能在膨胀相中存在。因为在收缩相那里没有强的热力学时间箭头，所以不适合智慧生命的存在。

人类理解宇宙的进步，在一个无序度增加的宇宙中建立了一个很小的有序的角落。如果你记住了这本书中的每一个词，你的记忆就记录了大约200万单位的信息——你头脑中的有序度就增加了大约200万单位。然而，当你读这本书时，你至少将以食物为形式的1千卡路里（1热化学卡＝4.184焦）的有序能量，以对流和出汗释放到你周围空气中的热量的形式，换成为无序能量。这就将宇宙的无序度增大了大约20亿亿亿单位，或大约是你头脑中有序度增量——那是如果你记住这本书的每一件事的话——的1000亿亿倍。我试图在下面各章再增加一些我们头脑的有序度，解释人们如何将我描述过的部分理论结合在一起，形成一个完备的统一理论，这个理论将适用于宇宙中的任何东西。

# 第十章 虫洞和时间旅行

我们在上一章讨论了，为什么我们看到时间向前进：为什么无序度增加，并且我们记住过去而非将来。时间好像是一条笔直的铁轨，人们只能往一个方向前进。

如果该铁轨有环圈以及分岔，使得一直往前开动的火车却返回原先通过的车站，这是怎么回事呢？人们能否旅行到未来或过去呢？

H·G·韦尔斯在《时间机器》中，正好像其他无数的科学幻想作家那样，探讨了这些可能性。科学幻想的许多观念，诸如潜水艇以及飞往月亮等都已变成了科学的事实。那么，时间旅行的前景如何呢？

1949年库尔特·哥德尔发现了广义相对论允许的新的时空。这首次表明物理学定律的确允许人们在时间里旅行。哥德尔是一名数学家，他由于证明了不完备性定理而名震天下。该定理是说，不可能证明所有真的陈述，即使你只试图证明像算术这么明确而且枯燥的学科中所有真的陈述。这个定理也许是我们理解和预言宇宙能力的基本极限，然而至少迄今为止，它似乎还未成为我们寻求完备统一理论的障碍。

哥德尔在和爱因斯坦于普林斯顿高级学术研究所度过他们晚年时通晓了广义相对论。他的时空具有一个古怪的性质：整个宇宙都在旋转。人们也许会问：“它相对于何物旋转？”其答案是远处的物体围绕着小陀螺或者陀螺仪的指向旋转。

这导致一个附加的效应，一位航天员可能在他乘航天飞船出发之前即已回到地球。这个性质使爱因斯坦非常沮丧，他曾经以为广义相对论不允许时间旅行。然而，鉴于爱因斯坦对引力坍缩和不确定性原理的无端反对，这也许反而是一个令人鼓舞的迹象。因为我们可以证明，我们生存其中的宇宙是不旋转的，所以哥德尔找到的解并不对应于它。它还有一个非零的宇宙常数。宇宙常数是当爱因斯坦以为宇宙是不变时引进的。在哈勃发现了宇宙的膨胀后，就不再需要宇宙常数，而现在普遍认为它应为零。然而，之后从广义相对论又找到其他一些更合理的时空，它们允许旅行到过去。其中之一即是旋转黑洞的内部。另外一种是包含两根快速相互穿越的宇宙弦的时空。顾名思义，宇宙弦是弦状的物体，它具有长度，但是截面很微小。实际上，它们更像在巨大张力下的橡皮筋，其张力大约为1亿亿亿吨。把一根宇宙弦系到地球上，就会把地球在1/30秒的时间里从每小时零英里加速到每小时60英里。宇宙弦初听起来像是科学幻想物，但有理由相信，在早期宇宙中由在第五章讨论过的那种对称破缺机制可以形成宇宙弦。因为宇宙弦具有巨大的张力，而且可以从任何形态起始，所以它们一旦伸展开来，就会加速到非常高的速度。

哥德尔解和宇宙弦时空一开始就这么扭曲，使得总能旅行到过去。上帝也许会创生了一个如此卷曲的宇宙，但是我们没有理由相信他会这样做。微波背景和轻元素丰度的观测表明，早期宇宙并没有允许时间旅行的曲率。如果无边界设想是正确的，从理论的基础上也能导出这个结论。这样问题就变成：如果宇宙初始就没有时间旅行必需的曲率，我们能否随后把时空的局部区域卷曲到这种程度，直至允许时间旅行？

快速恒星际或星系际旅行是一个密切相关的，也是科学幻想作家关心的问题。根据相对论，没有东西比光运动得更快。因此，如果我们向我们最近邻的恒星——半人马座α——发送航天飞船，由于它大约在4光年那么远，所以我们预料至少要8年才能等到旅行者们回来报

告他们的发现。如果要去银河系中心探险，至少要10万年才能返回。相对论确实给了我们一些宽慰。这就是在第二章提及的双生子佯谬。

因为时间不存在唯一的标准，而每一位观察者都拥有他自己的时间。这种时间是用他携带的时钟来测量的，这样航程对于空间旅行者比对于留在地球上的人显得更短暂是可能的。但是，这对于那些只老了几岁的返回的空间旅行者，并没有什么值得高兴的，因为他发现留在地球上的亲友们已经死去几千年了。这样，科学幻想作家为了使人们对他们的故事有兴趣，必须设想有朝一日我们能运动得比光还快。这些作家中的大部分并未意识到的是，如果你能运动得比光还快，相对论意味着，你就能向时间的过去运动，正如以下五行打油诗描写的那样：

有位年轻小姐名怀特，

她能行走得比光还快。

她以相对性的方式，

在当天刚刚出发，

却已在前晚到达。

关键在于相对论认为不存在让所有观察者同意的惟一的时间测量。相反，每位观察者各有自己的时间测量。如果一枚火箭能以低于光的速度从事件A（譬如2012年奥林匹克竞赛的100米决赛）至事件B（譬如半人马座α议会第100004届会议的开幕式），那么根据所有观察者的时间，他们都同意事件A发生于事件B之先。然而，假定飞船必须以超过光的速度旅行才能把竞赛的消息送到议会。那么，以不同速度运动的观察者关于事件A和事件B何为前何为后就众说纷纭。按照一位相对于地球静止的观察者的时间，议会开幕也许是在竞赛之后。这样，这位观察者会认为，如果他不理光速限制的话，该飞船能及时

地从A赶到B。然而，在半人马座α上以接近光速在离开地球方向飞行的观察者就会觉得事件B即议会开幕，先于事件A即百米决赛发生。相对论告诉我们，对于以不同速度运动的观察者，物理定律是完全相同的。

这已被实验很好地检验过。人们认为，即使用更高级的理论去取代相对论，它仍然会被作为一个特性保留下来。这样，如果超光速旅行是可能的，运动的观察者会说，就有可能从事件B，也就是议会开幕式，赶到事件A，也就是百米竞赛。如果他运动得更快一些，他甚至还来得及在赛事之前赶回，并在得知谁是赢家的情形下放下赌金。

要打破光速壁垒存在一些问题。相对论告诉我们，飞船的速度越接近光速，用以对它加速的火箭功率就必须越来越大。对此我们已有实验的证据，但不是航天飞船的经验，而是在诸如费米实验室或者欧洲核子研究中心的粒子加速器中的基本粒子的经验。我们可以把粒子加速到光速的99.99%，但是不管我们注入多少功率，也不可能把它们加速到超过光速壁垒。航天飞船的情形也是类似的：不管火箭有多大功率，也不可能加速到光速以上。

这样看来，快速空间旅行和逆时旅行似乎都不可行了。然而，还可能有办法。人们也许可以把时空卷曲起来，使得A和B之间有一近路。在A和B之间创生一个虫洞就是一个法子。顾名思义，虫洞就是一个时空细管，它能把两个相隔遥远的几乎平坦的区域连接起来。

虫洞两个端点之间在几乎平坦的背景里的分离和通过虫洞本身的距离之间没必要有什么关系。这样，人们可以想像，他可以创造或者找到一个从太阳系附近通到半人马座α的虫洞。虽然在通常的空间中地球和半人马座α相隔20万亿英里，而通过虫洞的距离却只有几百万英里。这就允许百米决赛的消息赶在议会开幕式前到达。然后一位往地球飞去的观察者也应该能找到另一个虫洞，使他从半人马座α议会开幕在赛事之前回到地球。这样，虫洞正和其他可能的超光速旅行方

式一样，允许人们逆时旅行。

时空不同区域之间的虫洞的思想并非科学幻想作家的发明，它的起源是非常令人尊敬的。

1935年爱因斯坦和纳珍·罗森写了一篇论文。在该论文中他们指出广义相对论允许他们称为“桥”，而现在称为虫洞的东西。爱因斯坦-罗森桥不能维持得足够久，使得航天飞船来得及穿越：虫洞会缩紧，而飞船会撞到奇点上去。然而，有人提出，一个先进的文明可能使虫洞维持开放。可以这样做，或者把时空以其他方式卷曲，使它允许时间旅行，人们可以证明，这需要一个负曲率的时空区域，如同一个马鞍面。通常的物质具有正能量密度，赋予时空以正曲率，如同一个球面。这样，为了使时空卷曲成允许逆时旅行的样子，人们需要负能量密度的物质。

能量有点像金钱：如果你有正的余额，就可以用不同方法分配，但是根据本世纪初（即20世纪初——编者注）相信的经典定律，你不允许透支。这样，这些经典定律排除了时间旅行的任何可能性。然而，正如在前面几章描述的，以不确定性原理为基础的量子定律已经超越了经典定律。量子定律更慷慨些，只要你总的余额是正的，你就允许从一个或两个账号透支。换言之，量子理论允许在一些地方的能量密度为负，只要它可由在其他地方的正的能量密度所补偿，使得总能量保持为正的。量子理论允许负能量密度的一个例子是所谓的卡西米尔效应。正如我们在第七章看到的，甚至我们认为是“空虚的”空间也充满了虚的粒子和反粒子对，它们一起出现相互分离，再返回一起，并且相互湮灭。现在，假定人们有两片距离很近的平行金属板，且金属板对于虚光子或光的粒子起着类似镜子的作用。事实上，在它们之间形成了一个空腔，它有点像风琴管，只对指定的音阶共鸣。这意味着，只有当平板间的距离是虚光子波长（相邻波峰之间的距离）的整数倍时，这些虚光子才会发生在平板之中的空间。如果空腔的宽度是波长的整数倍再加上部分波长，那么在前后反射多次后，一个波

的波峰就会和另一个的波谷重合，这样波动就被抵消了。

因为平板之间的虚光子只能具有共振的波长，所以虚光子的数目比在平板之外的区域要略少些，在平板之外的虚光子可以具有任意波长。这样，撞击在平板内表面的虚光子比外表面的略少一些。因此，人们可以预料到这两片平板遭受到把它们往里挤的力。实际上已经测量到这种力，并且和预言的值相符。这样，我们得到了虚粒子存在并具有实在效应的实验证据。

在平板之间存在更少虚光子的事实意味着，它们的能量密度比他处更小。但是在远离平板的“空虚的”空间的总能量密度必须为零，因为否则的话，能量密度会把空间卷曲起来，而不能保持几乎平坦。这样，如果平板间的能量密度比远处的能量密度更小，它就必须为负的。

这样，我们对以下两种现象都获得了实验的证据。第一，从日食时的光线偏折得知时空可以被卷曲。第二，从卡西米尔效应得知时空可被弯曲成允许时间旅行的样子。所以，人们也许希望，随着科学技术的推进，我们最终能够造出时间机器。但是，如果这样的话，为什么从来没有一个来自未来的人回来告诉我们如何实现呢？鉴于我们现在处于初级发展阶段，也许有充分理由认为，让我们分享时间旅行的秘密是不智的。除非人性得到彻底改变，非常难以相信，某位从未来飘然而至的访客会贸然泄漏天机。当然，有些人会宣称，观察到幽浮（UFO）就是外星人或者来自未来的人们来访的证据。（如果外星人在合理的时间内到达此地，他们则需要超光速旅行，这样两种可能性其实是等同的。）

然而，我认为，任何外星来的或者来自未来的人的造访应该是更明显得多，或许更加令人不悦得多。如果他们全然有意显灵的话，为何只对那些被认为不太可靠的证人进行？如果他们试图警告我们大难临头，这样做也不是非常有效的。

未曾有过对来自未来的访客，这可以用以下方法解释，因为我们观察了过去，并且发现它并没有允许从未来旅行返回必需的那类卷曲，所以过去是固定的。另一方面，未来是未知的开放的，所以也可能有需要的曲率。这意味着，任何时间旅行都被限制于未来。此时此刻，柯克船长和探险号星际飞船没有机会来临。

这也许可以解释，当今世界为何还没被来自未来的游客充斥。但是如果可能回到以前并改变历史，则不能够回避引起的问题。例如，假定你回到过去并且将你的曾曾祖父在他仍为孩童时杀死。这类佯谬有许多版本，但是它们根本上是等效的：如果一个人可以自由地改变过去，则他就会遇到矛盾。

看来有两种方法解决由时间旅行导致的佯谬。我把一种称为协调历史方法。它是讲，甚至当时空被卷曲得可能旅行到过去时，在时空中发生的必须是物理定律的协调的解。根据这个观点，除非历史表明，你曾经到达过去，并且当时并没有杀死你的曾曾祖父或者没有干过任何事和你的现状相冲突，你才能在时间中回到过去。此外，当你回到过去，你不能改变记载的历史。那表明你并没有自由意志为所欲为。当然，人们可以说，自由意志反正是虚幻的。如果确实存在一套制约万物的完整的统一理论，它也应该决定你的行动。但是对于像人这么复杂的机体，其制约和决定方式是不可能计算出来的。我们之所以说人们具有意志，乃在于我们不能预言他或她的未来行动。然而，如果一个人乘航天飞船出发并在这之前已经回返，我们就将能预言其未来行为，因为那将是记载的历史的一部分。这样，在这种情形下，时间旅行者没有自由意志。

解决时间旅行的其他可能的方法可称为选择历史假说。其思想是，当时间旅行者回到过去，他就进入和记载的历史不同的另外历史中去。这样，他们可以自由地行动，不受和原先的历史相一致的约束。史蒂芬·斯匹柏十分喜爱影片《回归未来》中的创意：玛提·马克弗莱能够回到过去，而且把他双亲恋爱的历史改得更令人满意。

听起来，选择历史假说和理查德·费恩曼把量子理论表达成历史求和的方法相类似，这已在第四章和第八章描述过。这是说宇宙不仅仅有一个单独历史，它有所有可能的历史，每一个历史都有自己的概率。然而，在费恩曼的设想和选择历史之间似乎存在一个重要的差别。在费恩曼求和中，每一个历史都是由完整的时空和其中的每一件东西组成的。时空可以被卷曲成可能乘火箭旅行到过去。但是火箭要留在同一时空即同一历史中，因而历史必须是协调的。这样，费恩曼的历史求和设想似乎支持协调历史假说，而不支持选择历史假说。

费恩曼历史求和确实允许在微观的尺度下旅行到过去。我们在第九章看到，科学定律在CPT联合作用下不变。这表明，一个在反时钟方向自旋并从A运动到B的反粒子还可以被认为是在时钟方向自旋并从B运动回A的通常粒子。类似地，一个在时间中向前运动的通常粒子等价于在时间中往后运动的反粒子。正如在本章以及第七章讨论过的，“空虚的”空间充满了虚的粒子和反粒子对，它们一道出现、分离，然后回到一块并且相互湮灭。

这样，人们可以把这对粒子认为是在时空中沿着一个闭合圈环运动的单独粒子。当对子在时间中向前运动时（从它出现的事件出发到达它湮灭的事件），它被称为粒子。但是，当粒子在时间中往回运动时（从对湮灭的事件出发到达它出现的事件），可以说成反粒子在时间中向前运动。

在解释黑洞如何发射粒子并辐射（见第七章）时认为，虚的粒子/反粒子对中的一个成员（譬如反粒子）会落到黑洞中去，另一个失去和它湮灭的伙伴的成员留了下来。这个被抛弃的粒子也可以落入黑洞，但是它也可以从黑洞的邻近挣脱。如果这样的话，对于一位远处的观察者，它就作为从黑洞发射出的粒子而出现。

然而，人们对于黑洞辐射的机制可有不同的却是等价的图象。人们可以把虚对中的那个落入黑洞的成员（譬如反粒子）看成从黑洞出

来的在时间中往回运动的粒子。当它到达虚粒子反粒子对一道出现的那一点，它被引力场散射成从黑洞逃脱的在时间中向前运动的粒子。相反，如果虚对中的粒子成员落入黑洞，人们也可以把它认为是从黑洞出来的在时间中往回运动的反粒子。这样，黑洞辐射表明，量子理论在微观尺度上允许在时间中的往回运动，而且这种时间旅行能产生可观测的效应。

因此产生这样的问题：量子理论在宏观尺度上允许时间旅行吗？这是人们能够利用的。初看起来，它应该是能够的。费恩曼历史求和的设想是指对所有的历史进行的。这样，它应包括被卷曲成允许旅行到过去的时空。那么，为什么我们并没有受到历史的骚扰？例如，假定有人回到过去，并把原子弹秘密提供给纳粹？

如果我称作时序防卫猜测成立的话，这些问题便可以避免。它是讲，物理学定律共谋防止宏观物体将信息传递到过去。它正如宇宙监督猜测一样，还未被证明，但是有理由相信它是成立的。

相信时序防卫有效的原因是，当时空被卷曲得可以旅行到过去时，在时空中的闭合圈环上运动的虚粒子，能够变成在时间前进的方向上以等于或者低于光速的速度运动的实粒子。由于这些粒子可以任意多次地围绕着圈环运动，它们通过路途中的每一点许多次。这样，它们的能量被再三地计入，使能量密度变得非常大。这也许赋予时空以正的曲率，因而不允许旅行到过去。这些粒子会引起正的还是负的曲率，或者由某种虚粒子产生的曲率是否被别种粒子产生的抵消，仍然不清楚。这样，时间旅行的可能性仍然未决。但是我不准备为之打赌，我的对手或许具有通晓未来的不公平的优势。

# 物理学的统一

正如在第一章中所解释的，一蹴而就地建立一个包括宇宙中万物的完备统一理论是非常困难的。取而代之，我们在寻求描述有限范围事件的部分理论上取得了进步，这时我们忽略了其他效应，或者将它们用一定的数字来近似表示（例如，化学允许我们计算原子间的相互作用时，可以不管原子核内部的结构）。然而，最终人们希望找到一个完备的协调的，将所有这些部分理论当作它的近似的统一理论。在这理论中不需要为某些任意数选值去符合事实。寻找这样的一个理论被称之为“物理学的统一”。爱因斯坦用他晚年的大部分时间寻求一个统一理论，但是没有成功。因为尽管已有了引力和电磁力的部分理论，但关于核力还知道得非常少，所以时机还没成熟。并且，尽管他本人对量子力学的发展起过重要的作用，但他拒绝相信它的真实性。但是，不确定性原理似乎还是我们生活其中的宇宙的一个基本特征。因此，一个成功的统一理论必须将这个原理结合进去。

正如我将描述的，由于我们对宇宙知道得这么多，现在找到这样的一个理论的前景似乎是好得多了。但是我们必须小心，不要过分自信——我们在过去有过对成功的错误的期望！例如，在本世纪初，曾经以为任何东西都可以按照诸如弹性和热传导之类的连续物质的性质予以解释。原子结构和不确定性原理的发现使之彻底破产。然后又有一次，1928年物理学家诺贝尔奖获得者马克斯·玻恩告诉一群来格丁根大学的访问者：“据我们所知，物理学将在6个月之内结束。”他的信心是基于狄拉克新近发现的能够制约电子的方程。人们认为质子——这个当时仅知的另一种粒子——服从类似的方程，并且那将会是

理论物理的终结。然而，中子和核力的发现对此又是当头一棒。尽管讲到这些，仍然有理由谨慎地乐观，我们现在也许已经接近探索自然终极定律的尾声。

在前几章中，我描述了引力的部分理论即广义相对论和制约弱力、强力和电磁力的部分理论。这后三种理论可以合并成为所谓的大统一理论（GUT）。这个理论并不令人非常满意，因为它没有包括引力，并且因为包含不能从理论预言，而必须人为选择以和观测符合的一些量，譬如不同粒子的相对质量，等等。要找到一个将引力和其他力统一的理论，主要困难在于广义相对论是一个“经典”理论；也就是说，它没有将量子力学的不确定性原理结合进去。另一方面，其他的部分理论却以非常基本的形式依赖于量子力学。因此，第一步必须将广义相对论和量子力学结合在一起。正如我们已经看到的，这能产生一些显著的推论，例如黑洞不是黑的，宇宙没有任何奇点，是完全自足的并且没有边界。正如第七章解释的，麻烦在于，不确定性原理意味着甚至“空虚的”空间也充满了虚的粒子和反粒子对，这些粒子对似应具有无限的能量，并且由爱因斯坦的著名方程$E=mc^2$得知，这些粒子似应具有无限的质量。这样，它们的引力的吸引就会将宇宙卷曲到无限小的尺度。

相当类似地，在其他部分理论中也发生似乎荒谬的无限大。但是，所有这些情形下的无限大都可用称作重正化的过程消除掉。这牵涉到引入其他的无限大去消除这些无限大。虽然这个技巧在数学上颇令人怀疑，而在实际上似乎确实行得通，并用来和这些理论一起作出预言，这些预言极其精确地和观测一致。然而，从企图找到一个完备理论的观点看，由于不能从理论中预言，而相反的为了适合观测，必须选择质量和力的强度的实际值，因此重正化确实具有一个严重的缺陷。

在试图将不确定性原理结合到广义相对论时，人们只有两个可以调整的量：引力强度和宇宙常数的值。但是调整它们不足以消除所有

的无限大。因此，人们得到一个理论，它似乎预言了诸如时空的曲率的某些量真的无限大，但是观察和测量表明它们地地道道是有限的！人们对这个结合广义相对论和不确定性原理的问题怀疑了许久，直到1972年才被仔细地计算最后确证。4年之后，人们提出了一种叫做“超引力”的可能的解答。它的思想是将携带引力的自旋为2称为引力子的粒子和某些其他具有自旋为3/2、1、1/2和0的新粒子结合在一起。在某种意义上，所有这些粒子可认为是同一“超粒子”的不同侧面。这样就将自旋为1/2和3/2的物质粒子和自旋为0、1和2的携带力的粒子统一起来了。自旋1/2和3/2的虚的粒子反粒子对具有负能量，因此抵消了自旋为2、1和0的虚的粒子对的正能量。这就使得许多可能的无限大被抵消掉，但是人们怀疑，可能仍然保留了某些无限大。人们需要找出是否还遗存下未被抵消的无限大。然而，这计算是如此之冗长和困难，以至于没人准备着手去进行。即使使用一个计算机，预料至少要用4年工夫，而且犯至少一个或更多错误的机会是非常大的。这样，只有其他人重复计算，并得到同样的答案，人们才能判断已取得了正确的答案，但这似乎是不太可能的！

尽管这些问题，尽管超引力理论中的粒子似乎与观察到的粒子不相符合的这一事实，大多数科学家仍然相信，超引力可能是对于物理学统一问题的正确答案。看来它是将引力和其他力相统一起来的最好办法。然而，1984年人们的看法发生了显著的改变，他们更喜欢所谓的弦理论。在这些理论中，基本的对象不再是只占空间单独的点的粒子，而是只有长度而没有其他维，像是一根无限细的弦这样的东西。这些弦可以有端点（所谓的开弦），或它们可以自身首尾相接成闭合的圈子（闭弦）。一个粒子在每一时刻占据空间的一点。这样，它的历史可以在时空中用一根线代表（“世界线”）。另一方面，一根弦在每一时刻占据空间的一根线。这样它在时空里的历史是一个叫做世界片的二维面。（在这世界片上的任一点都可用两个数来描述：一个指明时间，另一个指明这一点在弦上的位置）。一根开弦的世界片是一条带子：它的边缘代表弦的端点通过时空的路径（图11.1）。一根闭

弦的世界片是一个圆柱或一个管（图11.2）：一个管的截面是一个圈，它代表在一特定时刻的弦的位置。

两根弦可以连接在一起，形成一根单独的弦；在开弦的情形下只要将它们的端点连在一起即可（图11.3）。在闭弦的情形下，像是两条裤腿合并成一条裤子（图11.4）。类似地，一根单独的弦可以分成两根弦。在弦理论中，原先以为是粒子的东西，现在被描绘成在弦里旅行的波动，如同振动着的风筝的弦上的波动。一个粒子从另一个粒子发射出来或者被吸收，对应于弦的分解和合并。例如，太阳作用到地球上的引力，在粒子理论中被描述成由太阳上的粒子发射出并被地球上的粒子吸收的引力子（图11.5）。在弦理论中，这个过程对应于一个H形状的管（图11.6）（在某种方面，弦理论有点像管道工程）。H的两个垂直的边对应于太阳和地球上的粒子，而水平的横杠对应于在它们之间旅行的引力子。

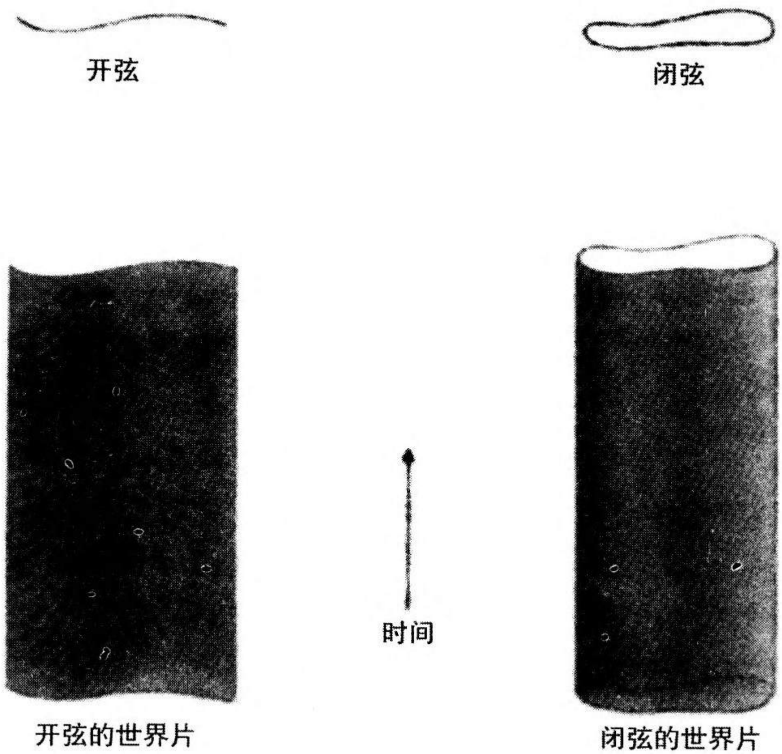



图11.1（左），图11.2（右）

弦理论有一个古怪的历史。它原先是60年代后期被发明出来，以试图找到一个描述强力的理论。其思想是，诸如质子和中子这样的粒子可被认为是一根弦上的波动。这些粒子之间的强力对应于连接其他一些弦之间的弦的片短，正如在蜘蛛网中一样。这弦必须像具有大约10吨拉力的橡皮带，才能使这理论给出粒子之间强力的观察值。

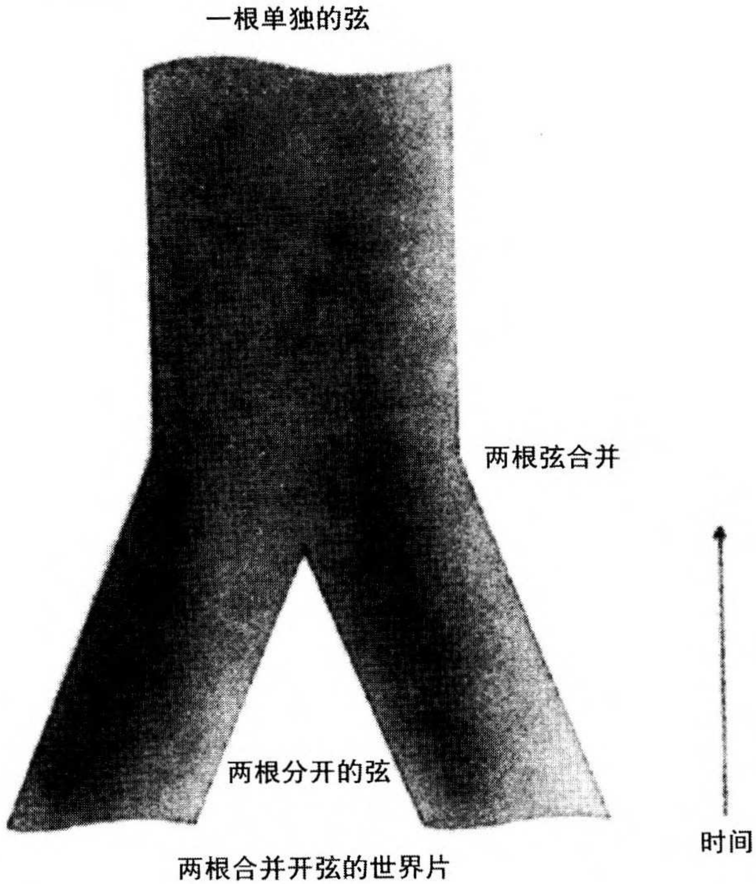



图11.3

1974年，巴黎的朱勒·谢尔克和加州理工学院的约翰·施瓦兹发表了一篇论文，指出弦理论可以描述引力，只不过其张力要大得多，大约是1000万亿亿亿亿吨（1后面跟39个0）。在通常尺度下，弦理论和广义相对论的预言是相同的，但在非常小的尺度下，比十亿亿亿亿分之一厘米（1厘米被1后面跟33个0除）更小时，它们就不一样了。然而，他们的工作并没有引起很大的注意，因为大约正是那时候，大多数人抛弃了原先的强作用力的弦理论，而倾心于基于夸克和胶子的理论，后者似乎和观测符合得更好。谢尔克死得很惨（他受糖尿病折

磨，在周围没人给他注射胰岛素时昏迷死去）。这样一来，施瓦兹几乎成为弦理论的惟一支持者，只不过现在设想的弦张力要大得多而已。

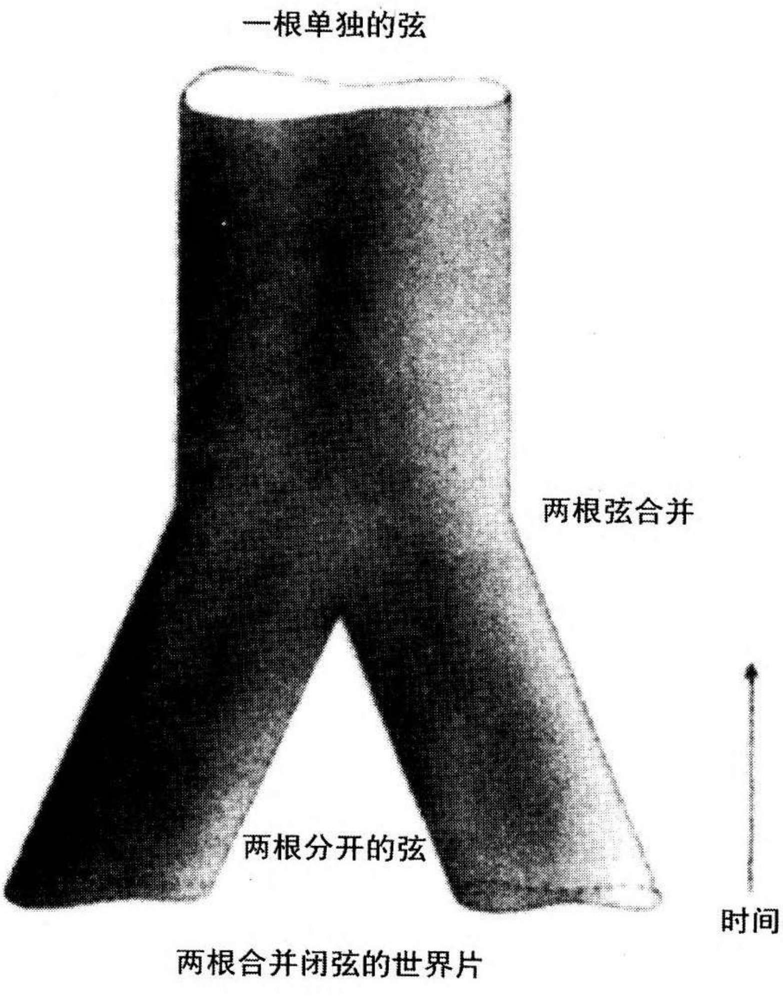



图11.4

1984年，由于两个明显的原因，人们对弦理论的兴趣突然复活。一个原因是，在证明超引力是有限的以及解释我们观察到的粒子的种

类方面，人们未能真正取得进展。另一个原因是，约翰·施瓦兹和伦敦玛丽皇后学院的迈克·格林发表的一篇论文指出，弦理论可以解释内禀的左手征性的粒子存在，正如我们观察到的一些粒子那样。不管什么原因，大量的人很快开始作弦理论的研究，而且发展了称之为杂化弦的新形式，这种形式似乎能够解释我们观测到的粒子类型。

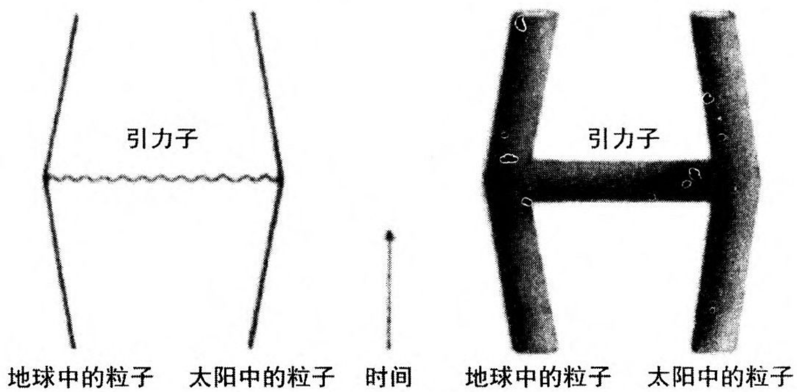



图11.5（左），图11.6（右）

弦理论也导致无限大，但是人们认为，它们在一些像杂化弦的形式中会被消除掉（虽然这一点还没被确认）。然而，弦理论有更大的问题：似乎时空是十维或二十六维，而不是通常的四维时它们才是协调的！当然，额外的时空维的确是科学幻想的老生常谈；它们提供了克服广义相对论的通常限制的理想方法，即人们不能行进得比光更快或者旅行到过去的限制（见第十章）。其思想是穿过更高的维抄近路。你可用以下方法描绘这一点。想像我们生活的空间只有二维，并且弯曲成像一个锚圈或环的表面（图11.7）。如果你处在这环的内侧的一边，而要跨过环到另一侧的一点去，你必须沿着环的内边缘上的圆圈走，直到目标点。然而，你如果允许在第三维空间里旅行，你可以直接穿过去。

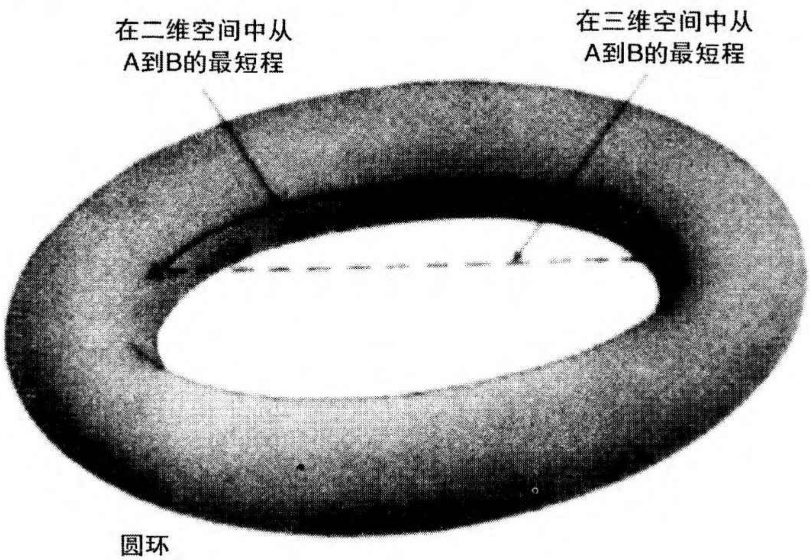



图11.7

如果这些额外的维确实存在，为什么我们全然没有觉察到它们呢？为何我们只看到三个空间维和一个时间维呢？人们的看法是，其他的维被弯卷到非常小的尺度——大约为一百万亿亿亿分之一英寸的空间，人们根本无从觉察这么小的尺度：我们只能看到一个时间维和三个空间维，在这些维中时空是相当平坦的。这正如一根麦秸的表面。如果你近看它，就会发现它是二维的（要用两个数来描述麦秸上的点，一个是沿着麦秸的长度，另一个是围绕着圆周方向的距离）。但是，当你远看它时，你分辨不出它的粗细，而它就显得是一维的（只用沿麦秸的长度来指明点的位置）。对于时空亦是如此：在非常小的尺度下，时空是十维的，并且是高度弯曲的，但是在更大的尺度下，你看不见曲率或者额外的维。如果这个图象是正确的，对于自愿的空间旅行者来讲可是个坏消息：额外的维实在是太小了，根本不允许航天飞船通过。然而，它引起了另一个重要问题。为何一些而非所有的维都被卷曲成一个小球？也许在宇宙的极早期，所有的维都曾经非常弯曲过。为何一维时间和三维空间被摊平开来，而且他维仍然紧

紧地卷曲着？

人存原理可能提供一个答案。二维空间似乎不足以允许像我们这样复杂生命的发展。例如，在一维地球上生活的二维动物，为了相互通过，就必须一个爬到另一个之上。如果二维动物吃东西时不能将之完全消化，则它必须将其残渣从吞下食物的同样通道吐出来，因为如果有一个贯通全身的通道，它就将这生物分割成两个分开的部分：我们的二维动物就解体了（图11.8）。类似地，在二维动物身上实现任何血液循环都是非常困难的。

多于三个空间维也有问题。两个物体之间的引力将随距离衰减得比在三维空间中更快（在三维空间内，如果距离加倍，则引力减少到1/4。在四维空间减少到1/8，五维空间1/16，等等）。其意义在于使像地球这样围绕着太阳的行星的轨道变得不稳定：地球偏离圆周轨道的最小微扰（例如由于其他行星的引力吸引）都会使它以螺旋线的轨道向外离开或向内落到太阳上去。我们就会被冻死或者被烧死。事实上，在维数多于三维的空间中，引力随距离变化的同样行为意味着，太阳不可能存在于压力和引力相平衡的稳定的状态下，它要么被四分五裂，要么坍缩形成一个黑洞。在任一种情况下，对地球上的生命来说，它作为热和光的来源都没有多大用处。在小尺度下，原子里使电子围绕着原子核运动的电力行为正和引力一样。这样，电子要么全部从原子逃逸出去，要么沿螺旋的轨道落到原子核上去。在任一种情形下，都不存在我们知道的原子。

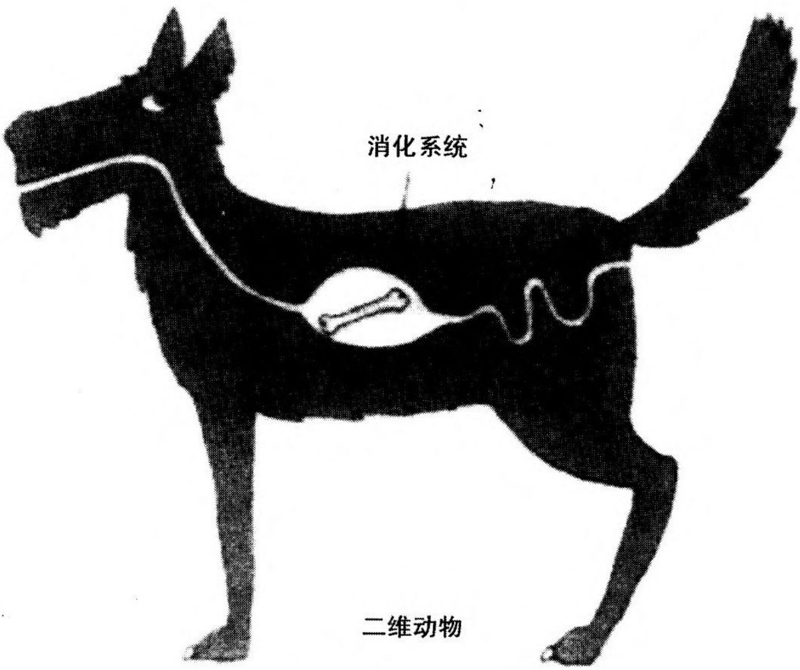



图11.8

看来很清楚，至少如我们所知，生命只能存在于一维时间和三维空间没被卷曲得很小的时空区域里。这表明，只要人们可以证明弦理论至少允许宇宙存在这样的区域——似乎弦理论确实能做到这一点，则我们可以求助弱人存原理。同样，也会存在宇宙的其他区域或其他宇宙（不管那是什么含意），那里所有的维都被卷曲得很小，或者多于四维几乎是平坦的，但是在这样的区域里，不会有智慧生物去观察这不同数目的有效维数。

另一个问题是至少存在四种不同的弦理论（开弦和三种不同的闭弦理论），以及由弦理论预言的额外维的极其繁多的卷曲方式。为何自然只挑选一种弦理论和一种卷曲方式？这问题一度似乎没有答案，因而无法向前进展。后来，大约从1994年开始，人们开始发现所谓的对偶性：不同的弦理论以及额外维的不同卷曲方式会导致四维时空中

的同样结果。不仅如此，正如在空间中占据单独一点的粒子，也像空间中线状的弦，还存在另外称作p膜的东西，它在空间中占据二维或更高维的体积。（粒子可认为是0膜，而弦为1膜，但是还存在p从2到9的p膜）。这似乎表明，在超引力、弦以及p膜理论中存在某种民主：它们似乎和平相处，没有一种比另一种更基本。看起来，它们是对某种基本理论的不同近似，这些近似在不同的情形下成立。

人们探索了这个基本理论，但是迄今毫无成就。正如哥德尔指出，不可能用单独的一组公理系统来表述算术。我相信这里的情形不可能比它更好，基本理论不可能存在单独的表述。相反，它也许和地图类似。你不可能只用一张单独的地图去描述地球或者锚圈的表面：在地球的情形下，你至少需要两张地图去覆盖每一点，而在锚圈的情形下，则需要四张。每张地图只对一个有限的区域有效，但是不同的地图有一个交叠的区域。整族地图就为该表面提供了完整的描述。类似地，在物理学中对不同的情形需要使用不同的表述，但是两种不同表述在它们都适用的情形下要相互一致。整族不同的表述可以被认为是完备的统一理论，尽管它不是依照单独的假设集合表达的理论。

但是，确实存在一个这样的统一理论吗？或者我们也许仅仅是在追求海市蜃楼。似乎存在三种可能性：

（1）确实存在一个完备的统一理论（或者一组交叠的表述），如果我们足够聪明的话，总有一天会找到它。

（2）并不存在宇宙的最终理论，仅仅存在一个越来越精确地描述宇宙的无限的理论序列。

（3）并不存在宇宙的理论：不可能在一定程度之外预言事件，事件仅以一种随机或任意的方式发生。

有些人基于以下理由赞同第三种可能，如果存在一组完备的定

律，这将侵犯上帝改变其主意并对世界进行干涉的自由。这有点像那古老的二律背反：上帝能制造一个重到连自己都不能将其举起的石块吗？但是上帝可能要改变主意的这一思想，正如圣·奥古斯丁指出的，是一个想像上帝存在在时间里的虚妄的例子：时间只是上帝创造的宇宙的一个性质。可以设想，当他创造宇宙时，他就知道了自己所有的企图！

随着量子力学的发现，我们认识到，由于总存在一定程度的不确定性，因此，不可能完全精确地预言事件。如果有人愿意，他可以将此随意性归结为上帝的干涉。但这会是一种非常奇怪的干涉：没有任何证据表明它具有任何目的。的确，如果它有目的，则按定义就不是随意的。现代由于我们重新定义科学的目标，所以已经有效地排除了上述的第三种可能性：我们的目的只在于表述一组定律，这些定律能使我们在不确定性原理的极限之内预言事件。

第二种可能性，也就是存在一无限的越来越精确的理论序列，是和迄今为止我们的经验相符合的。在许多场合我们增加了测量的灵敏度，或者进行了新的类型的观测，只是为了发现还没被现有理论预言的新现象，为了解释这些，我们必须发展更高级的理论。这一代的大统一理论预言：在大约100吉电子伏的弱电统一能量和大约1000万亿吉电子伏的大统一能量之间，没有什么本质上新的现象发生。因此，如果这个预言是错的话，人们并不会感到非常惊讶。我们的确可以预期发现一些新的比夸克和电子——这些我们目前以为是“基本”粒子——更基本的结构层次。

然而，引力似乎可以为这个“盒子套盒子”的序列设下极限。如果人们有一个具有比1000亿亿吉电子伏（1后面跟19个0）的所谓普朗克能量更高能量的粒子，它的质量就会集中到如此的程度，它就会脱离宇宙的其他部分，而形成一个小黑洞。这样看来，当我们往越来越高的能量去的时候，越来越精密的理论序列确实应当有某一极限，所以必须有宇宙的终极理论。当然，普朗克能量离开大约100吉电子伏

——目前在实验室中所能产生的最大的能量——非常远，我们不可能在可见的未来用粒子加速器填补其间的差距！然而，宇宙的极早期阶段是这样大的能量必定发生的舞台。我以为，早期宇宙的研究和数学协调性的要求，很有可能会导致当今我们周围的某些人在有生之年获得一个完备的统一理论。当然，这一切都是假定我们首先不使自身毁灭的前提下而言的。

如果我们确实发现了宇宙的终极理论，这意味着什么？正如在第一章中解释的，因为理论不能被证明，我们将永远不能肯定，我们是否确实找到了正确的理论。但是如果理论在数学上是协调的，并且总是给出与观察一致的预言，我们便可以适度地相信它是正确的。它将给人类理解宇宙的智力斗争长期而光辉的历史篇章打上一个休止符。但是，它还会变革常人对制约宇宙定律的理解。在牛顿时代，一个受教育的人至少可能在梗概上掌握整个人类知识。但从那以后，科学发展的节奏使之不再可能。因为理论总是被改变以解释新的观察结果，它们从未被消化或简化到使常人能够理解。你必须是一个专家，即使如此，你只能有望正确地掌握科学理论的一小部分。另外，其发展的速度如此之快，在中学和大学所学的总是有点过时。只有少数人可以跟得上知识快速进步的前沿，但他们必须贡献毕生的精力，并局限在一个小的领域里。其余的人对于正在进行的发展或者它们产生的激动只有很少的概念。70年前，如果爱丁顿的话是真的，那么只有两个人理解广义相对论。今天，成千上万的大学研究生能理解，并且几百万人至少熟悉这个思想。如果发现了一套完备的统一理论，以同样方法将其消化并简化，以及在学校里至少讲授其梗概，这只是时间的迟早问题。我们那时就都能够对制约宇宙，并对我们的存在负责的定律有所理解。

即使我们发现了一个完备的统一理论，由于两个原因，这并不表明我们能够一般地预言事件。第一是量子力学不确定性原理给我们的预言能力设立的限制。对此我们无法克服。然而，在实际上更为严厉

的是第二个限制。它是由以下事实引起的，除了非常简单的情形，我们不能准确解出这理论的方程。（在牛顿引力论中，我们甚至连三体运动问题都不能准确地解出，而且随着物体的数目和理论复杂性的增加，困难愈来愈大。）除了在最极端条件下之外，我们已经知道规范物体在所有条件下的行为的定律。特别是，我们已经知道作为所有化学和生物基础的基本定律。我们肯定还没有将这些学科归结为可解问题的状态：到现在为止，我们在根据数学方程来预言人类行为上只取得了很少的成功！所以，即使我们确实找到了基本定律的完备集合，在未来的岁月里，我们仍面临着在智慧上挑战性的任务，那就是发展更好的近似方法，使得在复杂而现实的情形下，能作出对可能结果的有用预言。一个完备的协调的统一理论只是第一步：我们的目标是完全理解发生在我们周围的事件以及我们自身的存在。

# 第十一章 结论

我们发现自己处于令人困惑的世界中。我们要理解周围所看到的一切的含义，并且寻问：宇宙的本质是什么？我们在其中的位置如何，以及宇宙和我们从何而来？宇宙为何是这个样子？

我们试图采用某种世界图来回答这些问题。如同一个无限的乌龟塔背负平坦的地球是这样的图象一样，超弦理论也是一种图象。虽然后者比前者更数学化，更准确得多，但两者都是宇宙的理论。两个理论都缺乏观测的证据：没人看到一个背负地球的巨龟，但也没有人看到超弦。然而，龟理论作为一个好的科学理论是不够格的，因为它预言了人会从世界的边缘掉下去。除非可以用它解释人们在百慕大三角消失的传说，否则这个理论和经验不一致！

最早在理论上描述和解释宇宙的企图牵涉到这样一种思想：具备人类情感的灵魂控制着事件和自然现象，它们的行为和人类非常相像，并且是不可预言的。这些灵魂栖息在自然物体，诸如河流、山岳以及包括太阳和月亮这样的天体之中。我们必须向它们祈祷并供奉，以保证土壤肥沃和四季循环。然而，我们逐渐注意到一些规律：太阳总是东升西落，而不管我们是否用牺牲向太阳神供奉。此外，太阳、月亮和行星沿着可事先被预言得相当准确的轨道穿越天穹。太阳和月亮仍然可以是神祇，只不过是服从严格定律的神祇。如果你不将为约书亚停止太阳运行之类的神话信以为真，则这一切显然是毫无例外的。

最初，只有在天文学和其他一些情形下，这些规律和定律才是显而易见的。然而，随着文明的发展，特别是近300年间，越来越多的规律和定律得到发现。这些定律的成功，使得拉普拉斯在19世纪初提出科学的决定论；也就是他提议的，有一组定律存在，只要给定宇宙在某一时刻的状态，这些定律就能精确决定宇宙的演化。

拉普拉斯的决定论在两个方面是不完整的：它没讲应该如何选择定律，也没指定宇宙的初始状态。这些都留给了上帝。上帝会选择让宇宙如何开始并要服从什么定律，但是一旦开始之后，他将不再干涉宇宙。事实上，上帝被局限于19世纪科学不能理解的领域里。

我们现在知道，拉普拉斯对决定论的希望，至少按照他所想的方式，是不能实现的。量子力学的不确定性原理意味着，某些成对的量，比如粒子的位置和速度，不能同时被完全精确地预言。量子力学通过一类量子理论来处理这种情形，在这些理论中粒子没有精确定义的位置和速度，而是由一个波来代表。这些量子理论给出了波随时间演化的定律，在这种意义上，它们是宿命的。于是，如果我们知道某一时刻的波，我们便可以将它在任一时刻推算出。只是当我们试图按照粒子的位置和速度对波做解释的时候，不可预见性的随机的要素才出现。但这也许是我们的错误：也许不存在粒子的位置和速度，只有波。只不过是我们企图将波硬套到我们关于位置和速度的先入为主的观念之上而已。由此导致的不协调乃是表面上不可预见性的原因。

事实上，我们已经将科学的任务重新定义为，发现能使我们在由不确定性原理设定的界限内预言事件的定律。然而，还存在如下问题：如何或者为何选取宇宙的定律和初始状态？

我在本书中特地突出制约引力的定律，因为正是引力使宇宙的大尺度结构成形，即使它是四类力中最弱的一种。引力定律和直到相当近代还为人深信的宇宙在时间中不变的观念不相协调：引力总是吸引，这一事实意味着，宇宙的演化方式两者必居其一，要么正在膨

胀，要么正在收缩。按照广义相对论，宇宙在过去某一时刻肯定有一个具有无限密度的状态，亦即大爆炸，这是时间的有效起始。类似地，如果整个宇宙坍缩，在将来必有另一个无限密度的状态，亦即大挤压，这是时间的终结。即使整个宇宙不坍缩，在任何坍缩形成黑洞的局部区域里都会有奇点。这些奇点正是任何落进黑洞的人的时间终点。在大爆炸时和其他奇点，所有定律都失效，所以上帝仍然有完全的自由去选择发生了什么以及宇宙如何开始。

当我们将量子力学和广义相对论结合，似乎产生了前所未有的新的可能性：空间和时间一起可以形成一个有限的四维的没有奇点或边界的空间，这正如地球的表面，但具有更多的维。看来这种思想能够解释宇宙间已观察到的许多特征，诸如它的大尺度一致性，还有包括星系、恒星甚至人类等在小尺度上对此均匀性的偏离。但是，如果宇宙是完全自足的，没有奇点或边界，并且由统一理论完全描述；那么就对上帝作为造物主的作用有深远的含义。

有一次爱因斯坦问道："在建造宇宙时上帝有多少选择性？"如果无边界假设是正确的，上帝就根本没有选择初始条件的自由。当然，上帝仍有选择宇宙所服从的定律的自由。然而，这也许实在并没有那么多选择性；很可能只有一个或数目很少的完备的统一理论，例如弦论，它们是自洽的，并且允许像人类那样复杂结构的存在，这些结构能够研究宇宙定律并询问上帝的本性。

即使只有一种可能的统一理论，那也只不过是一组规则和方程而已。是什么赋予这些方程以活力去制造一个为它们所描述的宇宙呢？通常的科学方法，即建立一个数学模型，不能回答为什么会有一个为此模型所描述的宇宙这个问题。为什么宇宙要找这么多存在的麻烦？难道统一理论如此咄咄逼人，以至于其自身之实现不可避免？或者它需要一个造物主，若是这样，它对宇宙还有其他效应吗？又是谁创造了它？

迄今为止，大部分科学家太忙于发展描述宇宙为何物的理论，以至于没工夫过问为什么。另一方面，以寻根究底为己任的哲学家跟不上科学理论的进步。在18世纪，哲学家把包括科学在内的整个人类知识当做他们的领域，并讨论诸如宇宙有无开端的问题。然而，在19和20世纪，对哲学家或除了少数专家以外的任何人来说，科学变得过于专业性和数学化了。哲学家把他们的质疑范围缩小到如此程度，以至于连维特根斯坦，这位20世纪最著名的哲学家都说道："哲学余下的任务仅是语言分析。"这是从亚里士多德到康德哲学的伟大传统的何等堕落啊！

如果我们确实发现了一个完备的理论，在主要的原理方面，它应该及时让所有人理解，而不仅仅让几个科学家理解。那时我们所有人，包括哲学家、科学家以及普普通通的人，都能参与讨论我们和宇宙为什么存在的问题。如果我们对此找到了答案，则将是人类理性的终极胜利——因为那时我们知道了上帝的精神。

# 阿尔伯特·爱因斯坦

爱因斯坦与核弹政治的瓜葛是众所周知的：他签署了那封著名的致富兰克林·罗斯福总统的信，说服美国认真考虑他的想法，并且他在战后致力于阻止核战争的爆发。但是，这些不仅仅是一位科学家被拖人政界的孤立行动。事实上，用爱因斯坦自己的话来说，他的一生“一半用于政治，一半用于方程”。

爱因斯坦最早从事政治活动是在第一次世界大战期间，当时他在柏林当教授。由于目睹草菅人命而不胜厌恶，他卷入了反战示威。他拥护非暴力反抗以及公开鼓励人民拒绝服兵役，因而不受他的同事们欢迎。后来，在战时，他又致力于调解和改善国际关系。这也使他不受欢迎，而且他的政治态度很快使他难以访问美国，甚至连讲学都有困难。

爱因斯坦第二个伟大的事业是犹太复国主义。虽然他在血统上是犹太人，但他拒绝接受《圣经》上关于上帝的说法。然而，在第一次世界大战之前和期间，他逐渐意识到反犹主义，这导致他参加犹太团体，而后成为一个直言不讳的犹太复国主义的拥护者。再度不受欢迎也未能阻止他发表自己的主张。他的理论开始受到攻击，甚至有人成立了一个反爱因斯坦的组织。有一个人因教唆他人去谋杀爱因斯坦而被定罪（却只罚了6美元）。但爱因斯坦是冷静的。当一本题为《100个反爱因斯坦的作家》的书出版时，他反驳道：“如果真是我错了的话，有一个人反对我就足够了！”

1933年，希特勒上台了。爱因斯坦正在美国，他宣布不再回德国。后来纳粹冲锋队查抄了他的房子，并没收了他的银行存款。一家柏林报纸的头条写道：“来自爱因斯坦的好消息——他不回来了。”面对着纳粹的威胁，爱因斯坦放弃了和平主义，由于担心德国科学家会制造核弹，他终于建议美国应该发展自己的核弹。但是，甚至在第一枚原子弹爆炸之前，他就曾经公开警告过核战争的危险，并提议对核武器进行国际控制。

终其一生，爱因斯坦致力于和平的努力可能成效甚微——肯定不受欢迎。然而，1952年他得到担任以色列总统的提议，他对犹太复国主义事业的畅言无忌的支持得到了充分的承认。但他谢绝了。他说他认为自己在政治上过于天真。可是，也许他真正的理由却并非如此，再次引用他自己的话：“方程对我而言更重要些，因为政治是为当前，而方程却是永恒的东西。”

# 伽利略·伽利雷

伽利略可能比任何其他人更有资格称为近代科学的奠基人。他与天主教会名闻遐迩的冲突对他的哲学是极重要的，因为伽利略是最早做出如下论断的人之一：人类有望理解世界如何行为，而且我们能通过观察现实世界来做到这一点。

伽利略很早就相信哥白尼理论（即行星围绕太阳公转），但只有当他发现了支持这一观念的证据后，才公开支持。他用意大利文写有关哥白尼理论的文章（没有用通常的学院式拉丁文），并且他的观点很快就广泛地传播到大学之外。这惹怒了亚里士多德派的教授们，他们联合起来反对他，并极力说服天主教会禁止哥白尼主义。

伽利略为此担心，他赶到罗马去向天主教会当局当面申诉。他争辩道，《圣经》并不试图告诉我们任何科学理论，而且通常都假定，在《圣经》和常识发生矛盾的地方，《圣经》是以讽喻的方式叙述的。

但是教会害怕这样的丑闻可能削弱它对新教的斗争，所以采取了镇压的手段。1616年，天主教会宣布哥白尼主义是“虚假的和错误的”，并命令伽利略再也不准“保卫或坚持”这一学说。伽利略勉强接受了。

1623年，伽利略的一位老友成为教皇。伽利略立即试图为1616年的判决翻案。他失败了，但他设法获得了准许，在两个前提下写一本讨论亚里士多德和哥白尼理论的书：他不能有倾向，同时要得出结

论，无论如何人不能确定世界是如何运行的，因为上帝会以人难以想象的方式来达到同样的效果，而人类不能限制上帝的万能。

这本题为《关于两大世界体系的对话》的书，于1632年在审查官的全力支持下完成并出版了——并且立刻被全欧洲欢呼为文学和哲学的杰作。不久教皇就意识到，人们把这本书视为拥护哥白尼主义的令人信服的论证，后悔允许该书出版。教皇指出，虽有审查官正式批准出版该书，但伽利略依然违背了1616年的禁令。他把伽利略带到宗教法庭，宣布对他终身软禁，并命令他公开放弃哥白尼主义。伽利略第二次被迫服从。

伽利略仍然是一个忠实的天主教徒，但是他对科学独立的信仰从未动摇过。1642年，即他逝世前4年，当时他仍然被软禁着，他第二本主要著作的手稿被偷运给一个荷兰的出版商。正是这本被称为《两种新科学》的书，甚至比支持哥白尼更进一步，成为现代物理学的发端。

# 艾萨克·牛顿

艾萨克·牛顿是一个不讨人喜欢的人，他和其他院士的关系声名狼藉，并在激烈的争吵中度过晚年的大部分时间。随着那部肯定是物理学有史以来最有影响的书——《自然哲学的数学原理》的出版，牛顿很快就成为名重一时的人物。他被任命为皇家学会主席，并成为第一个被授予爵位的科学家。

不久，牛顿就与皇家天文学家约翰·夫莱姆斯梯德发生冲突。他起初曾为牛顿《自然哲学的数学原理》一书提供急需的数据，但是他后来却扣压了牛顿需要的资料。牛顿是不许别人回答“不”字的，他自封为皇家天文台的大总管，然后强迫立即公布这些数据。最后，他指使夫莱姆斯梯德的冤家对头爱德蒙·哈雷夺取夫莱姆斯梯德的工作成果，并且准备出版。可是夫莱姆斯梯德告到法庭去，在最紧要关头赢得了法庭的判决——不得发行这部剽窃的著作。牛顿被激怒了，作为报复，他在《自然哲学的数学原理》后来的版本中系统地删除了所有来自夫莱姆斯梯德的引证。

他和德国哲学家哥特夫瑞德·莱布尼茨之间发生了更严重的争论。莱布尼茨和牛顿各自独立地发展了称作微积分的数学分支，它是大部分近代物理的基础。虽然现在我们知道，牛顿发现微积分要比莱布尼茨早若干年，可是他比莱布尼茨晚很久才出版他的著作。于是发生了关于谁是第一个发现者的大争吵，科学家们激烈地为双方做辩护。然而值得注意的是，大多数为牛顿辩护的文章均出自牛顿本人之手，虽然是以他朋友的名义出版！当争论日趋激烈时，莱布尼茨犯了向皇家

学会起诉来解决争端的错误。牛顿作为其主席，指定一个清一色的由牛顿的朋友组成的“公正的”委员会来审查此案！更有甚者，牛顿后来自己写了一个委员会报告，并让皇家学会将其出版，正式地谴责莱布尼茨的剽窃行为。即便如此，牛顿心犹未足，他又在皇家学会的杂志上写了一篇匿名的、关于该报告的回顾。据报道，莱布尼茨死后，牛顿扬言他为“伤透了莱布尼茨的心”而洋洋得意。

在这两次争吵期间，牛顿已经离开剑桥和学术界。在剑桥他曾积极从事反天主教政治，后来在议会中也很活跃。最终，作为酬报，他得到皇家造币厂厂长的肥差。在这里，他以社会上更能接受的方式，施展他那狡猾和刻薄的能耐，成功地导演了一场反对伪币的重大战役，甚至将几个人送上了绞刑架。

# 小辞典

绝对零度：所能达到的最低温度，在这温度下物体没有热能。

加速度：物体速度改变的速率。

人存原理：我们之所以看到宇宙是这个样子，是因为如果它不是这样的话，我们就不会在这里去观察它。

反粒子：每个类型的物质粒子都有相对应的反粒子。当一个粒子和它的反粒子碰撞时，两者就湮灭，只留下能量。

原子：通常物质的基本单元，是由很小的核子（包括质子和中子）以及围着它转动的电子构成。

大爆炸：宇宙开端的奇点。

大挤压：宇宙终结的奇点。

黑洞：时空的一个区域，因为那里的引力是如此之强，以至于任何东西，甚至光都不能从该处逃逸出来。

卡西米尔效应：在真空中两片平行的平坦金属板之间的吸引压力。这种压力是由平板之间空间中的虚粒子的数目比正常数目减小而引起的。

昌德拉塞卡极限：一个稳定的冷星的可能的最大质量的临界值。

比这质量更大的恒星，则会坍缩成一个黑洞。

能量守恒：关于能量（或它的等效质量）既不能产生也不能消灭的科学定律。

坐标：指定时空中一点的位置的一组数。

宇宙常数：爱因斯坦使用的一个数学手段，它赋予时空一个嵌入的膨胀倾向。

宇宙学：对整个宇宙的研究。

暗物质：在星系、星系团以及可能在星系团之间的物质，这种物质不能直接被观测到，但是可以由它的引力效应被检测到，宇宙中的质量多达90%可能处于暗物质的形式。

对偶性：导致相同的物理结果的，表面上不同的理论之间的对应。

爱因斯坦一罗森桥：连接两个黑洞的时空的细管。还请参见虫洞。

电荷：粒子的一个性质，由于这性质粒子排斥（或吸引）其他带有相同（或相反）符号电荷的粒子。

电磁力：在带电荷的粒子之间引起的力；它是四种基本力中第二强的力。

电子：带有负电荷并围绕着原子核转动的粒子。

弱电统一能量：大约为100吉电子伏的能量，在比这能量更大时，电磁力和弱力之间的差别消失。

基本粒子：被认为不可再分的粒子。

事件：由它的时间和位置所指明的在时空中的点。

事件视界：黑洞的边界。

不相容原理：两个相同的自旋为1/2的粒子（在不确定性原理设定的极限之内）不能同时具有相同的位置和速度。

场：某种充满空间和时间的东西，与它相反的是在一个时刻只在一点存在的粒子。

频率：对一个波而言，在1秒钟内完整循环的次数。

伽马射线：波长非常短的电磁射线，是由放射性衰变或由基本粒子碰撞产生的。

广义相对论：爱因斯坦基于如下思想的理论，即科学定律对所有的观察者，不管他们如何运动，都必须是相同的。它将引力解释成四维时空的曲率。

测地线：两点之间最短（或最长）的路径。

大统一能量：人们相信，在比这能量更大时，电磁力、弱力和强力之间的差别消失。

大统一理论（**GUT**）：一种统一电磁力、强力和弱力的理论。

虚时间：用虚数测量的时间。

光锥：时空中的面，在上面标出光通过一给定事件的可能方向。

光秒（光年）：光在1秒（1年）的时间里走过的距离。

磁场：引起磁力的场，现在和电场合并成电磁场。

质量：物体中物质的量；它的惯性或对加速的抵抗。

微波背景辐射：起源于炽热的早期宇宙的灼热的辐射，现在它受到如此大的红移，以至于不以光而以微波（波长为几厘米的射电波）的形式呈现出来。

裸奇点：不被黑洞围绕的时空奇点。

中微子：只受弱力和引力影响的极轻的粒子。

中子：一种和质子非常类似的但不带电荷的粒子，在大多数原子的核中大约一半的粒子是中子。

中子星：在超新星爆发后，有时一个恒星中心的物质的核坍缩成一团密集的中子，这种余下的冷的恒星称做中子星。

无边界条件：宇宙是有限的但是没有边界的思想。

核聚变：两个核碰撞并合并形成单独的更重的核的过程。

核：原子的中心部分，只由质子和中子构成。在核中强作用力将质子和中子束缚在一起。

粒子加速器：一种利用电磁铁能够对运动的带电粒子加速，并给它们更多能量的机器。

相位：对一个波，特定的时刻在它循环中的位置：种它是否在波峰、波谷或它们之间的某点的标度。

光子：光的一个量子。

普朗克量子原理：光（或任何其他经典的波）只能被发射或吸收分立的量子，其能量与它们频率成正比，和它们的波长成反比的思想。

正电子：电子的反粒子（带正电荷）。

太初黑洞：在极早期宇宙中产生的黑洞。

比例：“X”正比于“Y”表示当Y被乘以任何数时，X也如此；“X反比于Y”，表示当Y被乘以任何数时，X被那个数除。

质子：一种和中子非常类似的但带正电荷的粒子，在大多数原子的核中大约一半的粒子是质子。

量子力学：从普朗克量子原理和海森伯不确定性原理发展而来的理论。

脉冲量：发射出射电波规则脉冲的旋转中子星。

量子：波可被发射或吸收的不可分的单位。

量子色动力学：描述夸克和胶子相互作用的理论。

夸克：感受强作用力的（带电的）基本粒子。每一个质子和中子都由3个夸克组成。

雷达：利用脉冲射电波的单独脉冲到达目标并折回的时间间隔来测量对象位置的系统。

放射性：一种类型的原子核自动分裂成其他类型的原子核。

红移：由于多普勒效应，从离开我们而去的恒星发出的光线的红化。

奇点：时空中的一点，在该处时空曲率（或者一些其他的物理量）变得无限大。

奇点定理：这定理是说，在一定情形下奇点必须存在，特别是宇宙必须起始于一个奇点。

时空：四维的空间，上面的点是事件。

空间维：三维中的任何一维——也就是除了时间维外的任何一维。

狭义相对论：爱因斯坦的基于如下思想的理论，即科学定律在没有引力现象时，对所有进行自由运动的观察者，无论他们的运动速度如何，都必须相同。

谱：构成一个波的分频率。太阳谱的可见光部分可以在彩虹中看见。

自旋：相关于但不等同于日常的自转概念的基本粒子的内部性质。

稳态：不随时间变化的态：一个以固定速率自转的球是稳定的，因为即便它不是静止的，在任何时刻它看起来都是等同的。

弦论：物理学的理论，在该理论中粒子被描述成弦上的波。弦具有长度，但没有其他维。

强力：4种基本力中最强的，作用距离最短的一种力。它在质子和中子中将夸克束缚在一起，并将质子和中子束缚在一起形成原子核。

不确定性原理：海森伯表述的一个原理，该原理说人们永远不能

够精确地同时知道粒子的位置和速度；对其中的一个知道得越精确，则对另外一个就知道得越不精确。

虚粒子：在量子力学中，一种永远不能直接检测到的，但其存在确实具有可测量效应的粒子。

波/粒二象性：量子力学中的概念，认为在波和粒子之间没有区别；粒子有时可以像波一样行为，而波有时可以像粒子一样行为。

波长：在一个波中，两个相邻波谷或波峰之间的距离。

弱力：4种基本力中仅次于引力的第二弱的，作用距离非常短的一种力。它作用于所有物质粒子，而不作用于携带力的粒子。

重量：引力场作用在物体上的力。它和质量成正比，但又不同于质量。

白矮星：一种由电子之间不相容原理排斥力所支持的稳定的冷的恒星。

虫洞：连接宇宙的遥远区域的时空细管。虫洞还可以连接到平行或婴儿宇宙，并且能够提供时间旅行的可能性。

# 感谢

我在撰写本书时得到多人帮助。我的科学同仁毫无例外地激发我的灵感。在漫长的岁月里，我主要的合作者为罗杰·彭罗斯、罗伯特·格罗许、布兰登·卡特、乔治·埃里斯、盖瑞·吉朋斯、唐·佩奇和詹姆·哈特尔。他们有求必应，我非常感激他们，同时也非常感激我的学生们。

我的一名学生布里安·维特在准备初版时提供了许多帮助。矮脚鸡图书公司的编辑彼德·古查底还给我写下无数评语，使本书改善甚多。此外，我还想感谢安德鲁·杜恩，他为我修改了正文。

如果没有眼前的这台交流系统，本书就写不成。这套称作平衡器的软件是加利福尼亚兰卡斯特文字处理公司的瓦特·沃尔托兹捐赠的。我的语言合成器是加利福尼亚太阳谷的语音处理公司捐赠的。剑桥适用通讯公司的大卫·梅森把合成器和手提电脑安装在我的轮椅之上。利用这个系统，我现在比我失声之前能更好地与人交谈了。

在我撰写和修改此书的年代里，有过许多秘书和助手。对于秘书们，我应特别感谢的有朱迪·费拉、安·拉弗、劳拉·珍翠、谢瑞尔·比林顿和苏·梅西。我的助手为柯灵·威廉斯、大卫·托马斯、雷蒙·拉夫勒蒙、尼克·菲利普、安德鲁·杜恩、斯图瓦·詹米森、约纳逊·布连奇利、提蒙·汉特、赛蒙·基尔、琼·罗杰斯和汤姆·肯达尔。尽管我是残废的，但是他们、我的护士、合作者、朋友以及家人们使我的生命非常充实并能进行研究。

# 第一推动丛书书目

宇宙系列

《时间简史》

[英]史蒂芬·霍金著　吴忠超　许明贤译

《时间简史续编》

[英]史蒂芬·霍金著　吴忠超　胡小明译

《黑洞与时间弯曲》

[美]基普.S.索恩著　李泳译

《霍金讲演录》

[英]史蒂芬·霍金著　杜欣欣　吴忠超译

《时空本性》

[英]史蒂芬·霍金　罗杰·彭罗斯著　杜欣欣　吴忠超译

《果壳里的60年》

[英]史蒂芬·霍金等著　李泳译

《时空的未来》

[英]史蒂芬·霍金等著　李泳译

《千亿个太阳》

[德]鲁道夫·基彭哈恩著　沈良照　黄润乾译

《物理天文学前沿》

[英]F·霍伊尔　[印]】J·纳里卡著　何香涛　赵君亮译

生命系列

《惊人的假说》

[英]F·克里克著　汪云九等译

《生命是什么》

[奥]埃尔温·薛定谔著　罗来鸥　罗辽复译

物理系列

《宇宙的琴弦》

[美]B·格林著　李泳译

《上帝与新物理学》

[英]保罗·戴维斯著　徐培译

《时间之箭》

[英]彼得·柯文尼等著　江涛译

《终极理论之梦》

[美]S·温伯格著　李泳译

《时间、空间和万物》

[英]B·K·里德雷著　李泳译

《亚原子粒子的发现》

[美]S·温伯格著　杨建邺　肖明译

综合系列

《皇帝新脑》

[英]罗杰·彭罗斯著　许明贤　吴忠超译

《数学：确定性的丧失》

[美]M·克莱因著　李宏魁译

《四维旅行》

[英]R·L·普瓦德万著　胡凯衡　邹若竹译

《逻辑的引擎》

[美]马丁·戴维斯著　张卜天译

《莎士比亚、牛顿和贝多芬》

[美]S·钱德拉塞卡著　杨建邺　王晓明译

《未来50年》

[美]J·布洛克曼编　李泳译

# 黑洞与时间弯曲

# 总序

科学，特别是自然科学，最重要的目标之一，就是追寻科学本身的原动力，或曰追寻其第一推动。同时，科学的这种追求精神本身，又成为社会发展和人类进步的一种最基本的推动。

科学总是寻求发现和了解客观世界的新现象，研究和掌握新规律，总是在不懈地追求真理。科学是认真的、严谨的、实事求是的，同时，科学又是创造的。科学的最基本态度之一就是疑问，科学的最基本精神之一就是批判。

的确，科学活动，特别是自然科学活动，比较起其他的人类活动来，其最基本特征就是不断进步。哪怕在其他方面倒退的时候，科学却总是进步着，即使是缓慢而艰难的进步。这表明，自然科学活动中包含着人类的最进步因素。

正是在这个意义上，科学堪称为人类进步的“第一推动”。

科学教育，特别是自然科学的教育，是提高人们素质的重要因素，是现代教育的一个核心。科学教育不仅使人获得生活和工作所需的知识和技能，更重要的是使人获得科学思想、科学精神、科学态度以及科学方法的熏陶和培养，使人获得非生物本能的智慧，获得非与生俱来的灵魂。可以这样说，没有科学的“教育”，只是培养信仰，而不是教育。没有受过科学教育的人，只能称为受过训练，而非受过教育。

正是在这个意义上，科学堪称为使人进化为现代人的“第一推

动”。

近百年来，无数仁人智士意识到，强国富民再造中国离不开科学技术，他们为摆脱愚昧与无知做了艰苦卓绝的奋斗，中国的科学先贤们代代相传，不遗余力地为中国的进步献身于科学启蒙运动，以图完成国人的强国梦。然而应该说，这个目标远未达到。今日的中国需要新的科学启蒙，需要现代科学教育。只有全社会的人具备较高的科学素质，以科学的精神和思想、科学的态度和方法作为探讨和解决各类问题的共同基础和出发点，社会才能更好地向前发展和进步。因此，中国的进步离不开科学，是毋庸置疑的。

正是在这个意义上，似乎可以说，科学已被公认是中国进步所必不可少的推动。

然而，这并不意味着，科学的精神也同样地被公认和接受。虽然，科学已渗透到社会的各个领域和层面，科学的价值和地位也更高了。但是，毋庸讳言，在一定的范围内，或某些特定时候，人们只是承认“科学是有用的”，只停留在对科学所带来的后果的接受和承认，而不是对科学的原动力，科学的精神的接受和承认。此种现象的存在也是不能忽视的。

科学的精神之一，是它自身就是自身的“第一推动”。也就是说，科学活动在原则上是不隶属于服务于神学的，不隶属于服务于儒学的，科学活动在原则上也不隶属于服务于任何哲学的。科学是超越宗教差别的，超越民族差别的，超越党派差别的，超越文化的地域差别的，科学是普适的、独立的，它自身就是自身的主宰。

湖南科学技术出版社精选了一批关于科学思想和科学精神的世界名著，请有关学者译成中文出版，其目的就是为了传播科学的精神，科学的思想，特别是自然科学的精神和思想，从而起到倡导科学精神，推动科技发展，对全民进行新的科学启蒙和科学教育的作用，为

中国的进步做一点推动。丛书定名为《第一推动》，当然并非说其中每一册都是第一推动，但是可以肯定，蕴含在每一册中的科学的内容、观点、思想和精神，都会使你或多或少地更接近第一推动，或多或少地发现，自身如何成为自身的主宰。

《第一推动》丛书编委会

# 序

这本书讲的是我们在空间和时间观念上的革命及其重要结果，有些结果现在还不明朗。它也是一个迷人的故事，作者曾亲历过为认识黑洞这一也许是宇宙间最神秘事物的奋斗和成功。

过去人们常想，地球表面显然是平直的，它要么无限延伸，要么存在着边界，如果谁愚蠢地走得太远，他准会落下去。麦哲伦（Magellan）和其他环球旅行者的安全返回，令人们相信地球表面是弯曲的，自我封闭为一个球面；但人们仍然想当然地以为，这个球存在于欧几里得几何法则意义上的平直空间，即平行线永不相交的空间。然而，1915年，爱因斯坦提出一个理论，把空间和时间结合为一种叫“时空”的东西，它不是平直的，而是被其中的物质和能量弯曲（或卷曲）了。在我们邻近，时空几乎是平直的，在正常情况下不会出现曲率带来的差异。但在宇宙更远的地方，时空弯曲的某些结果甚至比爱因斯坦所认识的还要惊人。一个结果是，恒星可能在自身引力作用下坍缩，使周围空间发生弯曲从而将自己同宇宙其他部分分裂开来。爱因斯坦本人不相信会发生这样的坍缩，但大多数人都证明了，这是他的理论所不可避免的结果。

那些人如何证明这个结果，如何发现坍缩在空间留下的黑洞的奇异性质，正是本书的主题。这是一部活的科学发现的历史，作者是发现的参与者；在这点上，它很像DNA结构（它使我们认识了遗传密码）的发现者沃森（James Watson）写的《双螺旋》。不过，与DNA发现不同的是，这里没有能指导探索者的实验结果，倒是黑洞理论早就建立起来了，那时还没有任何它们确实存在的观测证据。我不

知道科学中还有没有别的例子，一个伟大的推论是完全依靠思想的基础而成功提出的。它说明了爱因斯坦理论的巨大威力和深远意义。

我们还有许多未知的事情。例如，落进黑洞的物体和信息会发生什么？它们会在宇宙其他地方或在另一个宇宙重新出现吗？我们能让空间和时间充分卷曲而回到过去吗？这些是我们为了认识宇宙而正在追寻的一部分问题，也许真有人能从未来回到现在，把答案告诉我们。

S·霍金

# 前言

本书是在严格的物理学原理基础上写的，并结合了高度的想象。作者试图超越人们目前的牢固知识而进入一个与我们地球的日常生活全然不同的物理世界。他的主要目的是考察黑洞的里里外外——黑洞质量大，引力场强，实物粒子和光都不可能像平常离开太阳那样从它逃逸出来。观测者从远处靠近这样的黑洞会遭遇哪些事情，是根据广义相对论在“强引力”作用下还没有经过检验的预言而描述的；超出这个范围进入所谓黑洞“视界”区域的悬想，则是靠一种特别的勇气，实际上就是特别的狂想，这在索恩和他的国际伙伴中是很多的，而且他们乐此不疲。这令人想起一位知名物理学家的妙语：“宇宙学家多犯错误，但少有怀疑。”读这本书的人应怀着两个目标：学一些我们物理宇宙中的尽管奇异却真实的可靠事实；欣赏那些我们还不那么有把握的奇思妙想。

作为开场白，我首先应该指出，爱因斯坦的广义相对论这一思辨科学的最伟大创造，不过是在四分之三世纪前才建立的。20年代初，它解释了水星运动与牛顿引力理论预言的偏离，后来又解释了哈勃（Hubble）和他的同事们在威尔逊山天文台观测到的遥远星云的红移。两次胜利以后，它沉默了几年，那时，多数物理学家的注意力都转向了量子物理学的诠释，转向了核物理、高能粒子物理和观测宇宙学的进步。

黑洞的概念在牛顿引力理论发现后不久，就以思辨方式提出来了。后来发现，通过适当修正，黑洞概念在相对论里也能找到自然的地位，不过，这需要我们将基本方程的解外推到极强的引力场——爱

因斯坦当时认为这样的外推过程是可疑的。然而，钱德拉塞卡（Chandrasekhar）在1930年指出，根据相对论，质量超过某一临界值（即所谓钱德拉塞卡极限）的星体在耗尽高温的核能源后，将坍缩成为我们现在所说的黑洞。大约在30年代后期，茨维基（Zwicky）、奥本海默（Oppenheimer）和他的同事分别推广了这一工作。他们证明，存在一个质量范围，在此范围内的星体不会坍缩成黑洞，而将形成一种由致密的中子堆积构成的状态，即所谓的中子星。不论哪种情况，星体核能耗尽时的内部挤压都会伴随一个相对短时间的巨大能量喷发，喷发的结果，就是我们在遥远星云和银河系中偶尔会看到的光亮的超新星。

这些研究在第二次世界大战时中断了，然而在50年代和60年代，科学家们又怀着新的兴趣和热情回到了它的实验和理论的前沿，取得了三大进展。第一，从核物理和高能物理研究获得的知识在宇宙学理论中找到了自然的位置，支持了通常所说的宇宙形成的“大爆炸”理论。现在，许多证据都支持这样的观点：我们的宇宙是从一点由紧密堆积的粒子所形成的原初热汤（一般称它为“火球”）膨胀而来的。这一原初事件大约发生在100到200亿年前。对这个假说最戏剧性的支持也许是发现了出现在原初爆炸后期的波的退化遗迹。

第二，我们确实观测到了茨维基和奥本海默小组预言的中子星，它们的行为也同理论预言的一样。这使我们完全相信，超新星是经历了“最后的引力坍缩”（大概可以这么说）的恒星。如果说中子星能存在于某一确定的质量范围，那么也有理由认为黑洞是质量更大的恒星的产物，不过我们承认，大量的观测证据都将是间接的；事实上，这类间接的证据现在已经很多了。

最后，广义相对论的有效性还得到了另外几方面证据的支持。它们包括太阳系中航天器和行星轨道的高精度测量和某些星系对外来光线的“透镜”作用。最近又发现了大质量双星系运动的能量损失，可能是它们发射引力波的结果，这是相对论的一个重要预言。这些发现，

不但使我们敢于相信广义相对论在黑洞附近的那些未经证实的预言，也为我们洞开了更广阔的想象空间。

几年前，联邦基金会（Commonwealth Fund）在M.E.Mathoney主席提议下，决定资助一项图书计划，邀请在不同领域工作的科学家们为受过教育的普通读者介绍他们的工作。索恩教授是其中的一位，很高兴本计划将他的书作为这一系列出版物的第九本。

推荐本书的联邦基金图书计划咨询委员会由下列成员组成：Lewis Thomas，医学博士，主任；Alexander G.Bearn，医学博士，副主任；Lynn Margulis，哲学博士；Maclyn McCarty，医学博士；Lady Medawar，Berton Roueche，Frederick Seiz，哲学博士；Otto Westphal，医学博士，出版者代表是W.W.Norton＆Company，Inc.的副主席兼编辑，Edwin Barber。

Frederick Seitz

# 导引

## 这本书讲什么，怎么读？

30年来，我一直在探索，为的是去认识爱因斯坦为后代留下的遗产——他的相对论和相对论关于宇宙的预言——去寻找相对论失败的地方，看它如何失败，会有什么来取代它。

在探索中，我穿过奇异事物的迷宫：黑洞、白矮星、中子星、奇点、引力波、虫洞、时间弯曲和时间机器；在探索中，我学会了认识论：什么让理论成为“好的”？什么样的“超原理”主宰着自然法则？为什么即使在技术还无力检验预言时，物理学家也会认为我们知道自以为知道的事情？在探索中，我明白了科学家的头脑是如何工作的，我看到了不同头脑之间的巨大差异（如霍金的与我的），我知道了为什么为了更真切地理解宇宙需要那么多不同类型的科学家以他们自己的方式工作。通过我们和遍布全球的几百名参与者的探索，我认识了科学的国际性特征，科学团体在不同社会中的组织形式，以及科学与政治潮流的相互纠缠，特别像苏美之间的竞争。

我想在这本书里与非科学家和不同领域的科学家分享我的这些感受。一条历史线索将相互关联的话题串在一起，那是我们为阐明爱因斯坦的精神遗产而奋斗的历史，也就是我们从遗产中发现那些奇异的黑洞、奇点、引力波、虫洞和时间弯曲的预言的历史。

书从序幕开始：我要讲一个科幻故事，它会很快把读者引向书中的物理学和天文学概念。有些读者可能会对故事感到沮丧，这些概念（黑洞和它的视界、虫洞、潮汐力、奇点和引力波）来得太快，几乎

没有解释。我告诉大家，不要管它，好好看故事，留下一点儿印象。每个概念在正文中还会以更令人愉快的方式重新提出来。读过全书后，你可以再回来看序幕，慢慢体会它的专业趣味。

书的主体（第1章到第14章）与序幕有完全不同的风味。中心线索是历史的，在历史的线索中交织着别的东西。我会用几页篇幅来追溯历史，然后离开历史讲一些题外话，然后又回到历史中来。这样，读者可以看到不同的思想观念如何精美地交织在一起，它们来自物理学、天体物理学、科学哲学、科学社会学和政治学。

有些物理学东西可以粗略看过，书后的物理学词汇表可能有一点帮助。

科学是公共的事业，形成我们宇宙观念的思想不是来自哪一个人或者某几个人，而是来自许多人的共同奋斗。因此，书中出现了许多人物，为帮助读者记住那些多次出现的人，书后列了一个“人物表”，对他们作了简单的介绍。

科学研究同人生一样，不同的人可能在同时探索许多相同的问题，一时的顿悟可能是几十年前的某些思想带来的灵感，而那些思想在几十年间却被忽略了。为了让读者能对此有些感受，本书将在时间里跳跃，从60年代跳到30年代，然后又回到70年代的历史主流。对这样的时间旅行感到眼花的读者，可以看书后的历史年表。

我不求历史学家要求的完整、准确和公正。如果要求完整，多数读者会跟我一样厌倦地将书扔到一边；如果要求更准确，这本书就会堆满公式而成为难啃的专著；尽管我追求公正，但一定存在偏见。我和我要讲的东西关系太近了，我个人从60年代到今天都在亲历它的发展，我最好的几个朋友从30年代起就身在其中了。我力图通过大量的对其他探索者的录音访问（见参考文献）和更多地描述其他人的工作（见致谢）来弥补可能的缺陷，然而肯定还留着某些偏见。

为帮助那些想更完整、更准确、更公正地了解历史的读者，我在书后的注释里列举了很多历史记述的来源，也请读者去参考探索者们为向别人阐释自己的发现而写的原始专业论文。注释里还对某些问题作了更准确（因而技术性也更强）的讨论，这些问题在正文里可能因为太简单化而容易使人误会。[1]

记忆是靠不住的。相同事件的不同经历者，对那些事件可能有不同的回忆和解释，这些分歧我都放在注释中了。在正文里，我只谈自己对事物的最终看法，就当它们是真的了。但愿真历史学家能原谅我，非历史学家会感谢我。

我的导师约翰·惠勒（John Wheeler，他也是本书的中心人物之一）在我成长为一名物理学家的过程中，喜欢问他的朋友，“关于这样那样的事物，你学到的最重要的一样东西是什么？”很少有问题能记得这么清楚。这本书断断续续写了15年（多数是“断”的时候），快写完时，我也在问自己跟约翰同样的问题：“你想让你的读者从书中学到的最重要的一样东西是什么？”

我的回答是：人类思想那令人惊奇的力量——在迷途中往返，在思想里跳跃——去认识宇宙的复杂，发现主宰它的基本定律的终极的单纯、精妙和壮丽。

[1] 哪些内容在后面有注释，作者并没有在正文里标记，而是在书后的注释部分指出它所在的页码和段落；为了更醒目，中译本在正文中的相应地方标记了数字，在后面的注释里则不再重复这些文字。——译者

# 序幕：黑洞之旅

读者在

一个科幻故事里

遭遇黑洞

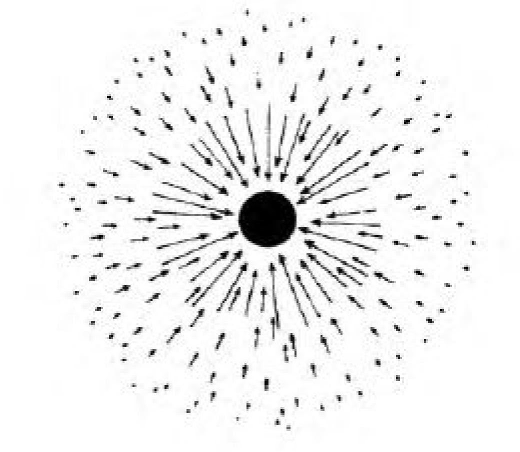
和我们在90年代所能认识的

关于它们的一切奇异性质

在人类头脑的所有概念中，[1]从独角兽到滴水嘴到氢弹，最奇异的也许还是黑洞：在空间中有一定边界的洞，任何事物都可以落进去，但没有东西能逃出来；一个强大引力能将光牢牢抓住的洞；一个能令空间弯曲和时间卷曲的洞。[1]跟独角兽和滴水嘴一样，黑洞似乎更多地出现在科幻小说和古代神话里，而不在真实的宇宙中。不过，经过了很好检验的物理学定律坚定地预言，黑洞是存在的，仅在我们的银河系里，可能就有几百万个，但它们太暗了，我们看不见；天文学家想发现它们也很困难。[2]

## 地狱

你有艘大飞船，自己做船长，带着计算机、机器人和几百名听话的船员，受世界地理学会委托，到遥远星际空间去探索黑洞，并把你的经历用电波发回地球。远航6年了，你的船正在减速接近织女星附近的一个黑洞，它叫“地狱”，离地球最近。[3]



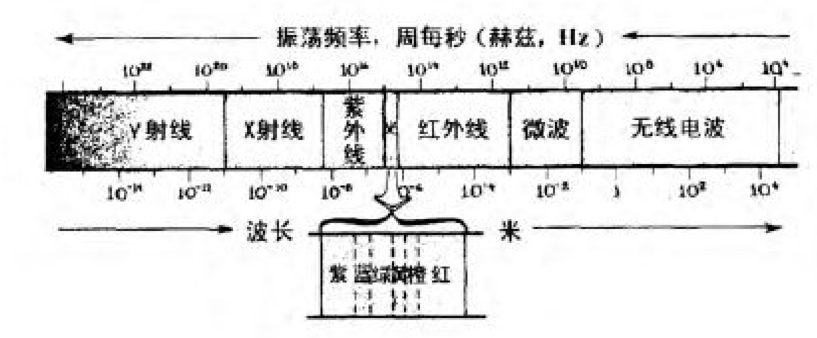
图P.1　在黑洞引力作用下，气体原子从各个方向流向黑洞。

你和船员从飞船的视屏上看到了黑洞出现的证据：散布在星际空间的气体原子（每立方厘米近1个）正受着黑洞引力的吸引（图P.1）。它们从所有方向流向黑洞，距离远的地方，引力作用较弱，原子流得较慢；距离近的地方，引力作用较强，原子流得较快——在靠近黑洞的地方，引力更强，原子流更快，几乎和光一样。假如不采取措施，飞船也会被黑洞吸进去。

大副卡丽丝迅速小心地将飞船从冲向黑洞的路线转到圆形轨道，然后关掉引擎。你们环绕着黑洞，飞船靠着圆周运动的离心力顶住了黑洞的引力。想想你小时候玩过的投石器，系在旋转绳子一端，离心力把它向外推，而绳子的张力将它往里拉；飞船像投石器，黑洞的引力就起着这种张力的作用。这时，你和船员准备开始探测黑洞。

先进行被动探测：用船上装备的望远镜研究电磁波（辐射），那是气体在流向黑洞时发射的。在远离黑洞的地方，气体原子很冷，只有绝对几度；因为冷，它们振动慢，缓慢的振动产生缓慢振荡的电磁波，意味着从一个波峰到下一个波峰的距离（即波长）很长。这些就是无线电波，见图P.2。在离黑洞较近的地方，引力作用下的原子流

较快，它们相互碰撞，加热到几千度的高温；因为热，它们振动较快，发出振荡较快、波长较短的波，也就是你所认识的不同颜色的光：红、橙、黄、绿、蓝、紫（图P.2）。离黑洞更近的地方，引力更强，原子流更快，碰撞更剧烈，温度更高（几百万度），原子极快地振动，产生波长很短的电磁波：X射线。看到从黑洞附近喷出的X射线时，你会想起，在1972年，天体物理学家就是因为发现和研究了这样的X射线，才认定了遥远空间的第一个黑洞：天鹅X-1，距地球14000光年。[4]




图P.2　电磁波谱，从波长很长（频率很低）的无线电波到波长很短（频率很高）的γ射线。图中所用数字记号（$10^{20}$，$10^{-12}$等）见后面卡片P.1的讨论。

把望远镜对准离黑洞更近的地方，你看到从被加热到更高温度的原子发射出的γ射线。接着，你看到，在这片辉煌的景象中心，出现了一个绝对黑暗的圆球，那是一个黑洞，吞噬了从它背后的原子所发出的一切可见光、X射线和γ射线。你注视着超热的原子从四处流进黑洞。一旦进了黑洞，它们会比以前更热，振动也一定比以前更快，辐射也会更强，但它们的辐射逃不脱黑洞强大的引力。没有什么东西能逃出来。这就是为什么那洞是黑的，漆黑的一团。[5]

你拿望远镜更真切地审视那个黑球，发现它有绝对分明的边缘，即黑洞的表面，一个“逃不脱”的地方。刚好在表面以上的东西，如果有足够的力量，可以逃脱引力的魔掌：火箭能飞走；向上发射足够快

的粒子能逃走；光当然也能逃走。但如果刚好在表面以下，那么不论是火箭、粒子、光、辐射或者其他任何东西，不论费多大力气，都不可能逃脱引力那无情的魔掌，永远不能到达你旋转的飞船。于是，黑洞的表面就像我们的地平线，你看不到它下面的东西。这也就是为什么我们把这表面称为黑洞的地平线。[6]

大副卡丽丝仔细测量了飞船轨道的周长，100万千米，大约是月亮绕地球轨道的一半。然后，她看外面遥远的恒星，看着它们在飞船头上旋转。通过测量恒星这种视运动的时间，她推测，飞船绕黑洞一周需要5分46秒，这就是飞船的轨道周期。

现在，你可以根据轨道周期和周长计算黑洞的质量。[2]计算方法和牛顿（Issac Newton）1685年计算太阳质量的方法相同：天体（太阳或黑洞）质量越大，它的引力作用越强，于是围绕它的物体（行星或飞船）为避免被它吸进去，必然也运动得越快，从而轨道周期一定就越短。用牛顿引力定律[7]的数学公式，你算出黑洞“地狱”的质量比太阳大10倍（“10个太阳质量”）。[8]

你知道，这个黑洞是很久以前恒星死亡形成的。恒星在死亡时顶不住自身引力的吸引而发生坍缩，就产生黑洞。[9]你也知道，恒星坍缩时质量不会改变，“地狱”今天的质量与它的母星很久以前的质量是一样的——或者说，几乎是一样的。实际上，自黑洞诞生以来，落进去的事物，如星际气体、岩石、飞船……都会增加它的质量，所以“地狱”的质量一定会比原来的母星大一点。

你知道这些，是因为你在旅行前学过引力的基本定律：牛顿在1687年发现了它的近似形式，爱因斯坦在1915年又从根本上修正了牛顿的近似，得到了更精确的形式。[10]你知道，黑洞的这些行为像石头落回地球一样，都是所谓广义相对论的爱因斯坦引力定律所要求的。石头不可能违背引力定律而向上落或者漂浮在天空，同样，黑洞也躲不开引力：它必然在恒星坍缩中诞生；它初生时的质量一定与恒

星质量相同；每次落进来的事物都一定会增大它的质量。[11]同样，假如坍缩恒星是旋转的，那么新生的黑洞也一定旋转；而黑洞的角动量（旋转快慢的精确度量）也一定与恒星的相同。

你在远航前还学过人类认识黑洞的历史。早在70年代，卡特尔（Brandon Carter）、霍金（Stephan Hawking）、伊斯雷尔（Werner Israel）和其他一些人就用爱因斯坦广义相对论表述的引力定律[12]发现，黑洞一定是极其简单的怪物：[13]黑洞的一切性质——它的引力作用强度、它对星光轨道的偏转、它的表面形状和大小——仅由三个参数决定：黑洞的质量，你已经知道了；黑洞旋转的角动量，你还不知道；还有黑洞的电荷。而且你还知道，星际空间的黑洞都不能带太多的电荷；假如电荷太多，它会很快从星际气体中吸引相反的电荷来中和自己的电荷。

黑洞旋转的时候，会像飞机旋转的螺旋桨带动空气那样，带着它附近的空间（相对于遥远的空间）像龙卷风一样作涡旋运动；空间的旋涡又在黑洞附近一切事物的运动中激起旋涡。[14]

于是，为了解“地狱”的角动量，你在落向黑洞的星际气体原子流中寻找龙卷风式的旋涡。你惊讶地发现，原子流离黑洞越来越近，运动越来越快，却没有任何旋涡的迹象。原子盘旋着落下，有些是顺时针的，另一些则是反时针的，它们偶尔会发生碰撞，但总的说来，是无旋涡地径直向着黑洞下落的。你认定：这个10个太阳质量的黑洞几乎没有旋转，它的角动量近乎零。

知道了黑洞的质量和角动量，又知道它的电荷一定少得可以忽略，现在你可以用广义相对论公式来计算黑洞应该具有的一切性质了：引力作用强度、相应的偏转星光的能力以及更有意义的，黑洞视界的形状和大小。

假如黑洞在旋转，视界会有分明的北极和南极，也就是黑洞旋转

的极点和下落的原子绕着它盘旋的极点。两极中间还会有明显的赤道，因视界旋转的离心力而向外凸起，跟旋转的地球赤道的凸起是一样的。[15]但“地狱”几乎没有旋转，所以一定不会有赤道的凸起。它的视界在引力作用下几乎完全是球形的，这正是你在望远镜里看到的样子。

至于大小，[3]广义相对论描述的物理学定律认为，黑洞质量越大，它的视界也一定越大。实际上，视界周长必然是以太阳质量为单位的黑洞质量乘以18.5公里。[16]你从轨道周长的测量得知黑洞有10个太阳那么重，因此视界周长肯定是185公里——和洛杉矶差不多大。你用望远镜仔细测量了周长，真是185公里，完全符合广义相对论的公式。

同你那100万公里的飞船轨道相比，视界的周长真是太小了，而被挤进这样一个小空间里的质量却有10个太阳那么大！假如黑洞是固体的，那么挤在这么小的空间里，它的平均密度将是每立方厘米2亿（$2\times10^8$）吨——比水重$2\times10^{14}$倍（参见卡片P.1）。但黑洞不是固体。广义相对论认为，10个太阳的星体物质在很久以前通过坍缩形成黑洞，现在聚集在黑洞的中心——聚集在一个叫做奇点的小空间区域里。[17]约$10^{-33}$厘米大小的奇点（比原子核小1万亿倍）周围，除了正在向它落下的稀薄气体和气体发出的辐射以外，什么也没有。从奇点到视界几乎是空虚的，从视界到你的飞船，也差不多是空的。

卡片P.1　大数和小数的幂表示

在本书中，我偶尔会用“幂记号”来表示很大和很小的数。例如，$5\times10^6$的意思是500万，或5000000，而$5\times10^{-6}$则是百万分之五，或0.000 005。

一般说，将幂表示的数变成标准的十进制数，就是将10的幂次作为小数点移动的位数。这样，$5\times10^6$意味着5（5 000 000 00）的小

数点右移6位，结果是5 000000.00。同样，$5\times10^{-6}$意味着5的小数点左移6位，结果是0.000 005。

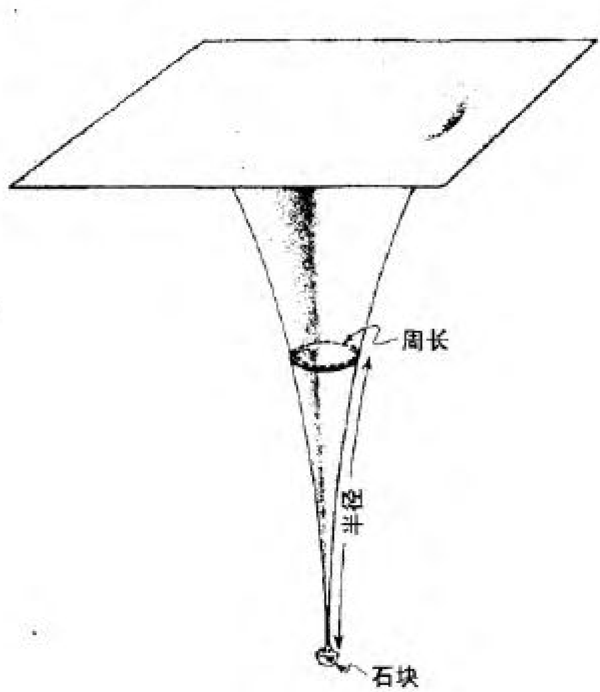
奇点和困在其中的星体物质躲在黑洞视界里，不论你等多久，被困的物质都不会再出现，黑洞的引力把它锁住了；它也不可能通过电波、光或者X射线向你传送信息。实际上，它完全从我们的宇宙消失了。惟一留下的是它强大的引力，对你那100万公里的轨道来说，它今天的引力作用与它在坍缩成黑洞以前的作用是一样的；但在视界内部，却没有什么东西能够抵抗它的引力了。

“视界距奇点多远呢？”你问自己。（你当然不会去测量它，那简直就是自杀；你也不可能从视界逃出来向世界地理学会报告你的测量结果。）由于奇点很小，只有$10^{-33}$厘米，正好在黑洞中心，所以从奇点到视界的距离应等于视界的半径。你忍不住想用标准的方法来计算半径：用周长除以2π（6.283 185 307…）。但是，你在地球上的研究中知道要警惕这样的计算，不能随便相信。黑洞的巨大引力彻底扭曲了黑洞内部和附近的空间和时间的几何，[18]仿佛放在一张橡皮上的沉重的石块扭曲了橡皮的几何（图P.3）；结果，视界的半径不等于它的周长除以2π。

“那没关系，”你告诉自己，“罗巴切夫斯基（Lobachevsky）、黎曼（Riemami）和其他伟大的数学家已经教过我们，如何在空间弯曲时计算圆的性质，爱因斯坦又把这些计算融人了他的引力定律的广义相对论描述。我可以用这些弯曲空间的公式来计算视界的半径。”

但是，你这时又想起，根据在地球上的研究，尽管黑洞的质量和角动量决定了视界和它外面的所有性质，但并没决定它的内部。广义相对论认为，在奇点附近，黑洞的内部应该是混沌的，绝不是球形的，[19]就像图P.3，一块棱角尖利的岩块重重地落在一张橡皮上，猛烈地弹起又落下，砸出一个尖尖的深坑。另外，黑洞中心的混沌性质不仅依赖于黑洞的质量和角动量，而且依赖于产生它的星体坍缩的细节

和后来落进的星际气体的细节——那是你还不知道的。




图P.3　一块重石头放在一张橡皮上使它变形。橡皮扭曲的几何类似于黑洞周围和内部空间变形的几何。例如，粗黑圆圈的周长远小于2π乘以它的半径，正如黑洞视界周长远小于2π乘以半径一样。进一步的讨论，见第3，13章。

“那又怎么样呢？”你对自己说，“不管混沌的黑洞中心有什么结构，它的周长总是远比1厘米还小的，这样，即使把它忽略了，我计算的视界半径也不会有太大的误差。”

然而，这时你又想起，空间在奇点附近可以极端卷曲，这样，混沌区域可能在不足1厘米的周长下有几百万千米的半径，就像图P.3的那块重重的石头把橡皮混沌的尖端砸得远远的，而区域的周长却是短短的。你的半径计算就可能会产生这么大的误差。视界的半径不可能简单地凭你掌握的那点儿黑洞质量和角动量的信息来计算。

你不再去想黑洞内部了，而准备探测它的视界附近。你不愿人拿生命去冒险，而让一个机器人去，并要他把探测结果传回飞船。一个10厘米高的机器人阿诺尔德（Arnold）将带着火箭去探险。他要做的事情很简单：先发动火箭，让自己从跟飞船一起的环行中停下来，然后关闭引擎，在黑洞引力作用下径直落下去。在下落中，阿诺尔德向飞船发出明亮的绿色激光束，光束的电磁振荡载着他下落距离和他的电子系统状态的信息，就像电台发射的无线电波载着广播新闻的信号。

船员收到发回的激光束后，卡丽丝将解译阿诺尔德的距离和系统的信息，并测量光束的波长（或者等价地说，测量它的颜色；见图P.2）。波长的重要，在于它能说明阿诺尔德的运动有多快。当他离开飞船的运动越来越快时，飞船收到的他传回的绿光会因为多普勒频移而显出越来越大的波长，越来越红的颜色。[20]（另外，还有部分由于光束摆脱引力作用而产生的红移。计算了阿诺尔德的速度后，卡丽丝会通过修正得到引力红移。）[21]

实验就这样开始了。阿诺尔德点燃火箭，离开飞船轨道，进入向黑洞下落的轨迹。在他开始下落时，卡丽丝开始计时，测量激光信号的到达时间。10秒过去了，激光信号表明一切系统运行正常，他已经下落了2 630千米。卡丽丝根据激光颜色算出，他现在正以每秒530千米的速度冲向黑洞。时钟走到20秒时，他下落的距离已经是刚才的4倍，10 520千米。时钟继续嘀嗒，60秒时，他的速度是每秒9 700千米，下落了135 000千米，到视界的距离过了5/6。

现在你必须密切注意了，接下来的几秒是决定性的。于是，卡丽丝打开高速记录系统来收集数据的所有细节。61秒，阿诺尔德报告，一切系统运行正常；视界在他下方14 000千米，他正以每秒13 000千米的速度落下去。61.7秒，仍然一切正常，还有1700千米了，速度是每秒39 000千米，约光速的1/10；激光颜色开始剧变。接着的1/10秒里，你惊奇地看到激光从电磁波谱匆匆掠过，从绿到红，到红外，

到微波，到无线电波——61.8秒时，它走完了，激光束完全消失了。阿诺尔德达到了光速，消失在视界里。在激光消失的最后1/10秒，阿诺尔德还在高兴地报告，“一切系统正常，正常；视界临近了，系统正常，正常……”

从激动中镇静下来，你开始来检验记录的数据。你看到了激光波长移动的整个过程。当阿诺尔德下落时，激光信号的波长先慢慢增长，然后越来越快。但令人惊讶的是，波长增到4倍后，它加倍的速率几乎是一个常数，即每0.000 14秒增加1倍。经过33次加倍（0.004 6秒）后，波长达到4千米，是你记录系统的极限。以后，波长大概还会加倍的。波长变得无限大，需经过无限次的加倍，所以，在黑洞视界邻近也许还会出现波长极大、极暗淡的信号！

这是不是说阿诺尔德还没穿过视界，而且永远不会穿过呢？不。那最后的一丝波长永远在加倍的信号需要无限长的时间才能从黑洞引力束缚中逃出来。阿诺尔德在很多分钟以前就以光速飞过了视界。那些还在继续出来的微弱信号不过是因为走得太慢了，它们是过去遗留下来的。[22]

你研究了很久阿诺尔德发回的数据，然后好好睡一觉，恢复了精神，准备做下一次探险。这回，你要亲自去视界邻近看看，而且比阿诺尔德仔细得多。

告别船员，你钻进太空舱，脱离飞船，进入它的圆形轨道。然后，你轻轻发动火箭，将轨道运动减慢，这也稍稍减弱了太空舱所受的离心力，而黑洞的引力将你拉到一个小一些的圆形轨道。接着，你再轻轻发动火箭，圆轨道会再收缩一点。你就想这样安全平稳地螺旋式地达到视界上方的轨道，它的周长刚好是视界周长的1.000　1倍。在这里，你可以探测视界的许多性质，还能逃脱它那要命的魔掌。

然而，在你的轨道慢慢收缩时，一些奇怪的事情也开始发生了。

在100　　000公里周长的轨道上，你就能感觉到它们。你漂在太空舱里，脚朝黑洞，头朝星空。你会感到有一股微弱的力量在把你的脚向下拉，而把你的头向上拉，就像拉一块太妃糖，不过力量小一些。你知道，那是黑洞引力的结果：脚比头离黑洞更近，所以它受到的黑洞引力作用比头更强。这在地球上当然也是对的；不过，在地球上头脚引力差别很小，不到百万分之一，你根本觉察不出来。而漂浮在100　000千米周长轨道的太空舱里，情况就不同了，头脚引力差别是地球引力的八分之一（1/8“g”）。在身体中心，轨道运动的离心力正好抵消黑洞引力，仿佛引力不存在，而你在自由漂浮着。但是，你的脚多受着l/16g的向下拉的引力，在你的头上，引力较弱，而向外推的离心力却多余1/16g。

你虽然感到惊讶，还是继续盘旋着下去；但是，你很快又忧虑起来。随着轨道缩小，头和脚的力量越来越强。在80　　000千米的轨道上，拉力是1/4g；50　000千米时，等于地球引力；30000千米时，是4倍地球引力。你咬牙忍着头脚分离的痛苦，继续下到20　　000千米的轨道，那儿力量是15g，再大你就忍不住了！你想把身体蜷缩起来，让头脚靠得近些，这样拉力可以小一点，但现在的拉力太强了，你不可能缩成一团，它总会在轨道半径方向上将你的头脚拉直。如果太空舱再落下去，你的身体就完了，会被完全撕裂！你没有希望到达黑洞的邻近。

你带着巨大的失望和痛苦停了下来，调转头，开始小心翼翼地回来。你盘旋着上升，穿过越来越长的轨道，最后回到飞船的货舱。

走进船长室，你就在主计算机DAWN上发泄你的失望。“提克哈依（Tikhii），提克哈依，”机器安慰你（用的是古俄语的词儿），“我知道你难过，但那都是你自己的错。在训练时就告诉过你那种头脚拉力的事儿，记得吗？它们就是地球上引起海洋潮汐的那种力。”[23]

你想起来了。你学过，在地球离月亮最近的一端，海洋受到最强的月亮引力，所以会涌向月亮。在相对的一端，海洋受的引力最弱，仿佛要离开月亮。结果，海洋在地球两端涌起，随地球自转，每24小时出现两次高潮。你记起来了，你经历的那种从头到脚的引力，就是这样的力，所以叫潮汐力。[4]你还记得，爱因斯坦的广义相对论把潮汐力描述为空间曲率和时间卷曲的结果，或者，用爱因斯坦自己的话说，是时空曲率。[24]潮汐引力与时空扭曲是并存的，一个总伴着另一个；不过，在海洋潮汐中，时空的扭曲太小，只有用极精确的仪器才能测量。

那么，阿诺尔德呢？他为什么一点儿也不怕黑洞的潮汐引力？DAWN解释说，原因有两点，第一，他比你小得多，只有10厘米高，作用在头和脚的引力差别相应也很小；第二，他是用超强钛合金做的，比你的骨头硬得多。

现在你明白了，阿诺尔德经历了多么可怕的一幕。当他穿越视界继续落向奇点时，一定感到潮汐力在增强，甚至最后超过了他那超强钛合金的抵抗能力。穿过黑洞0.000 2秒后，他破碎的身体接近了黑洞中心的奇点。这时，你又回忆起在地球上从广义相对论学到的东西：在那儿，黑洞的潮汐力又活跃起来了，混沌地跳跃着，在不同的方向拉扯阿诺尔德的残骸，一会儿这个方向，一会儿那个方向；越来越快，越来越强，最后他的每个原子都被扭曲而不能识别了。实际上，这就是奇点的本性：它是混沌振荡的时空曲率产生巨大随机潮汐力的一个区域。[25]

回忆黑洞研究的历史，你想起来了，1965年，英国物理学家彭罗斯（Roger Penrose）用广义相对论形式的物理学定律证明了奇点一定藏在黑洞内部；1969年，俄罗斯的栗弗席兹（Lifshitz）、卡拉特尼科夫（Khalatnikov）和别林斯基（Belinsky）这“三驾马车”发现，在奇点邻近，潮汐引力一定会混沌地振荡，它的行为就像我们做太妃糖，一会儿这么拉，一会儿那么压。[26]

六七十年代，黑洞理论研究的黄金年代啊！但是，黄金年代的物理学家还不能充分认识爱因斯坦的广义相对论方程，黑洞行为的一个关键特性还困扰着他们。他们只能猜想，坍缩的恒星不论什么时候产生奇点，总会产生包围隐藏奇点的视界；奇点不可能是“裸露”的，不会让全宇宙都看到它。彭罗斯称它为“宇宙监督猜想”，因为假如它是对的，那么它将监督所有关于奇点的实验信息：人们永远也不可能用实验来检验他们关于奇点的认识，除非谁愿意付出生命的代价走进黑洞去测量；即使那样，他还是不能把结果从黑洞传出来，连一点儿纪念物也不会留下。

也许，2023年会有某个叫奈曼（Dame Abygaole Lyman）的人能最终解决宇宙监督是否正确的问题，但那结果与你无关。你的地图上画的只是黑洞里的奇点，而你不愿为它们去死。

幸运的是，在接近黑洞视界的外面，仍然有许多可以探测的现象。你决定亲自去经历这些现象，然后向世界地理学会报告。但你不能到“地狱”视界的附近去，那儿的潮汐力太强了；你一定要找一个潮汐力弱一些的黑洞。

DAWN提醒你，广义相对论预言，黑洞质量越大，视界上和视界外的潮汐引力越弱。这个似乎矛盾的行为有很简单的原因：潮汐力正比于黑洞质量除以周长的立方；质量增加时，视界周长也正比例地增加，视界附近的潮汐力实际上减小了。[5]一个100万太阳质量的黑洞，也就是比“地狱”重100 000倍的黑洞，视界也将大100 000倍，它的潮汐力将弱100亿（$10^{10}$）倍。这是令人满意的，一点儿痛苦也不会有了！于是，你开始计划下一步的航行：去谢切特（Schechter）黑洞图上最近的那个100万太阳质量的黑洞——它叫“人马”（Sogittario），在银河系的中心，离我们30 100光年。

几天后，船员把“地狱”探险的报告，你被潮汐力拉伤的图像和原子落进黑洞的图像，都传回地球。26光年的距离，要走26年；报告最

后到达地球后，世界地理学会将大肆宣扬。

在发回来的报告里，船员还谈了你们去银河中心的远航计划：飞船的火箭将一路保持地球的重力加速度（lg），这样你和船员在飞船里就处在舒适的地球重力作用下。在前一半旅程里，飞船加速向银河中心，然后调转180°，以lg的加速度减速经历另一半旅程。整个旅程30 100光年，在地球看来，需要经过30 102年；但在飞船看来，只需要20年。[6]这是因为，根据爱因斯坦的狭义相对论定律，[27]高速的飞船会使飞船测量的时间"膨胀"；这种时间膨胀（或时间卷曲）在效果上就使飞船成了一台时间机器，让你在短暂的时间里走到地球遥远的未来。[28]

你们告诉世界地理学会，下一次消息将在探测了100万太阳质量的黑洞"人马"后，从银河系中心发回来。如果学会的会员想活着收到信息，他就得"冬眠"60 186年（从收到你们的消息到你们到达银河中心的时间是30 102-26＝30 076年；另外，你们下一次消息从银河中心传到地球还需要30 110年）。

[1] 第3，6，7章。

[2] 第8章。

[3] 织女星距地球26光年，飞船6年就能到它的邻近，为什么？请读者想想。——译者

[4] 第8章。

[5] 第3，6章。

[6] 第6章。（"黑洞的地平线"，在中文的物理学文献里一般称"视界"，以后我们也说"视界"。——译者）

[7] 第2章。

[8] 想自己计算黑洞性质的读者可以在书后的注释里找到相关的公式。

[9] 第3～5章。

[10]　第2章。

[11]　关于物理学定律强迫黑洞，太阳系和宇宙以一定方式行动的进一步讨论，请看第1章最后几段。

[12]　第2章。

[13]　第7章。

[14]　第7章。

[15]　第7章。

[16]　第13章。18.5公里这个量本书还会出现多次，它是4π（即12.5663706…）乘以牛顿引力常数乘以太阳质量，除以光速的平方。关于它和其他描述黑洞的公式，请看这一章的注释。

[17]　第13章。

[18]　第3，13章。

[19]　第13章。

[20]　见卡片2.3。

[21]　第2，3章。

[22]　第6章。

[23]　第2章。（著名的《静静的顿河》的俄文名字就是Tikhii Don.——译者）

[24]　第2章。

[25]　第13章。

[26]　第13章。

[27]　见第1章。

[28]　见第1章。

# 人马

经过20年的航行，飞船减速飞进了银河系中心。你远远看见气体和尘埃混合着从四面八方流向一个巨大的黑洞。卡丽丝调好火箭，将飞船带入视界上方的圆形轨道。你测量了轨道的周长和周期，把结果代进牛顿的公式，确定了黑洞的质量：100万个太阳质量，与谢切特黑洞图上说的一样。由于没有在下落的气体和尘埃里看到龙卷风似的旋涡，你推测黑洞不会旋转太快；视界应该是球状的，而周长一定是1 850万千米，是月球环绕地球轨道的8倍。

进一步检查下落气体后，你准备向视界靠近。为了安全，卡丽丝在你的太空舱和飞船主机DAWN间建立了激光联系。然后，你脱离飞船，调转太空舱，让它的喷气对着飞船轨道运动的方向；接着轻轻启动，使你的轨道运动慢下来，平稳地螺旋式地向里（向下）经过一个又一个圆形轨道。

一切都如预料的那样正常。但等到周长为5 500万千米的轨道——刚好是黑洞视界周长的3倍，火箭的推动却没有将你引入一个更小的轨道，而是要命地将你投向视界。你害怕极了，赶紧调转方向，以最大力量冲出来，回到5 500万千米以外的轨道。

“究竟出了什么事儿？！”你通过激光问DAWN。

“提克哈依，提克哈依，”她安慰你说，“你的轨道是根据牛顿的引力定律设计的，但牛顿的描述只是宇宙真实引力定律的一种近似。[1]在远离视界的地方，它是很好的近似，但在视界附近，它却糟透了。更精确的描述是爱因斯坦的广义相对论，在视界附近，它能以很高的精度与真实的引力定律一致。它预言，在接近视界时，引力作用会变得比牛顿预言的更强。为了保持圆形轨道，以离心力对抗强大的引力，你必须加强离心力，也就是说，你必须提高围绕黑洞的轨道速

度：当你下落经过3倍视界周长的轨道时，你必须调转太空舱的方向，向前加速；如果你还向后减速的话，在你经过那个轨道时，引力将超过离心力，把你拉下去。”[7]

“该死的DAWN！”你想，“她总能回答我的问题，却从来不主动提出一些关键信息。我要犯错误时她从不警告！”你当然知道这是为什么。假如计算机都在我们犯错误之前提出警告，人类生活该是多么枯燥乏味！早在2032年，世界委员会就通过法案，在每台计算机里都植入霍布森障碍，[2]不许计算机警告。尽管DAWN也许很愿意警告你，但她实在不能克服霍布森障碍。

你压住怒火，调转太空舱，开始一系列的操作：向前加速，向下盘旋，进入内轨道；再向前，再盘旋，进入下一个轨道……从3个视界周长降到2.5，2.0，1.6，1.55，1.51，1.505到1.501到……太令人失望了！你越飞越快，轨道越来越小；在飞行速度接近光速时，你的轨道却只到1.5个视界周长。因为不能比光速更快，看来你没有希望靠这种办法走近视界了。

你又向DAWN求救，她一样安慰你，然后解释，1.5个视界周长以内根本没有圆形轨道。那儿的引力作用太强，没有离心力可以同它对抗，哪怕你以光速绕着黑洞旋转。DAWN告诉你，如果想走得更近，你必须放弃圆形飞行轨道，而应该直接朝视界落下去。靠火箭向下喷气，你可以避免灾难性的坠落。火箭的反冲力可以为你克服一些引力，让你慢慢落下，然后漂浮在视界上方，就像宇航员漂浮在月球上飞行的火箭里。

现在，你学会小心了。你问DAWN，这样持续强烈的火箭喷射会有什么后果？你解释说，你想漂在1.000 1视界周长的某个位置，在那儿，能经历视界的多数效应，而且还可以逃出来。“如果凭火箭支持太空舱，那么加速度的力量会有多大呢？”“1.5亿地球引力。”DAWN轻轻回答。

真令人泄气！你点燃火箭，盘旋着飞回了飞船。

好好睡一觉。醒来后，你拿广义相对论的黑洞公式算了5个小时，在谢切特黑洞图上找了3个小时，又与船员们讨论了1个小时，最后确立了下一步航行计划。

接着，船员把你在“人马”的经历传给世界地理学会（你们乐观地假定它还存在着）。报告最后讲了你的计划：

你的计算表明，黑洞越大，你漂浮在1.000 1视界周长上所需要的火箭动力越小。[8]为了不超过10个地球引力（这虽然也痛苦，但还能忍受），黑洞必须有15万亿（$1.5\times10^{13}$）个太阳质量。最近的这样的黑洞叫“巨人”，远在距我们10万（$10^5$）光年的银河系以外，也远在银河系围绕的1亿（$10^8$）光年的室女座星系团以外；实际上，它在类星体3C273附近，距银河系20亿（$2\times10^9$）光年，大约是我们可以观测的宇宙边缘距离的10%。

船员在报告中解释，你的计划就是去“巨人”。前一半旅程以1g加速，后一半以1g减速，这样，在地球看来，旅行需要20亿年，而幸好因为有了速度产生的时间卷曲，你们在船上只需42年就够了。[9]如果世界地理学会不愿40亿年的漫长冬眠（飞船到“巨人”20亿年，信息发回地球20亿年），他们就收不到你们的下一次消息了。

[1] 见第2章。

[2] 16-17世纪英国租马房经营者霍布森（T.Hobson）规定租马不允许挑选；英语成语有Hobson's choice，指毫无选择余地的状态。这里说Hobson's block，大概意思是不可避免的阻碍。——译者

# 巨人

42年后，飞船减速来到“巨人”的邻近。你们的头上是类星体3C273，两股灿烂的蓝色喷流正从它的中心射出；[1]下面就是“巨人”那黑暗的无底洞。落在“巨人”外的一个轨道上，你做了些常规测量，证实了它的确具有15万亿个太阳质量；另外，它旋转很慢。从这些数据，你算出它的视界周长是29光年。现在，你终于找到你向往的黑洞了！你能到它的邻近去探险，而不会遭遇难以忍受的巨大潮汐力和火箭加速度。既然探险有了安全保障，你决定飞船整体下降，不再只凭一个太空舱。不过，在飞船下降前，你命令船员拍摄一些照片：头顶巨大的类星体，“巨人”周围数万亿颗恒星，天空中几十亿个星系。他们还拍了在下面的“巨人”的黑洞圆盘，像地球看到的太阳那么大。乍看起来，黑洞似乎挡住了所有来自它背后的恒星和星系的光；但仔细看时，你的船员发现，黑洞的引力场像一个透镜，[2]恒星和星系的光偏转绕过视界的边缘，在黑洞圆盘边缘又被聚焦成一条明亮的细环。每一颗朦胧的恒星在环上都有几个像，一个是从黑洞左翼绕过的光线产生的；另一个是从右翼绕过的光线产生的；第3个是被吸引到绕黑洞的轨道的光在你的方向上发出时形成的；第4个是绕黑洞两周后跑出来的光线形成的，等等。结果，光线形成一条结构高度复杂的环，为了将来的研究，船员们拍摄了大量的细节照片。

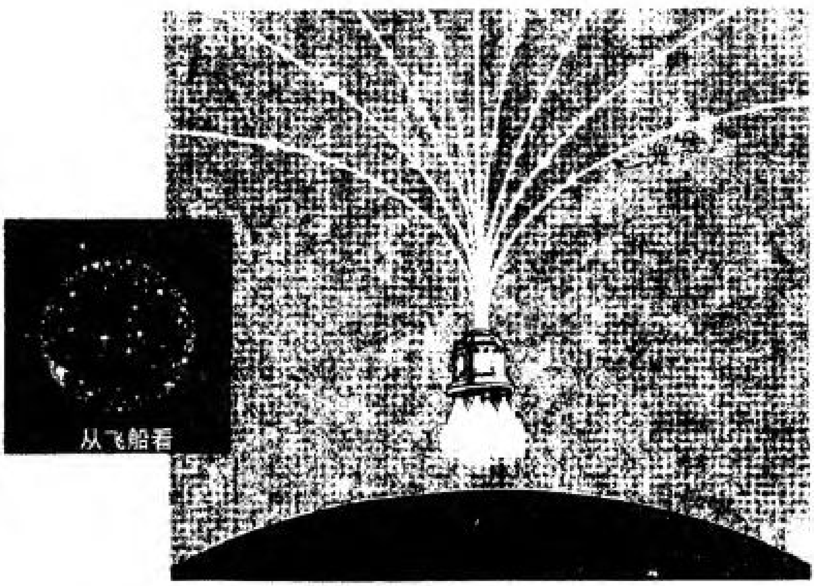
照片拍好后，你命令卡丽丝开始启动飞船降落。但是，你还得耐心一点，黑洞引力太强，你们以1g加速、减速，需要13年才能到达你们计划的1.000 1视界周长！

飞船落下来了，船员们又拍了些照片，记录飞船周围天空的变化。最引人注目的变化是，飞船下面的黑洞圆盘长大了：慢慢地越来越大。你想，它会像巨大的黑色地板铺满你的脚下，然后停下来，头上还是像地球上明朗的天空。但黑盘子仍然在长大，从飞船周围升

起，遮盖了一切，只留下头上一道明亮的圆形光路，你能从它看到外面的宇宙（图P.4）。你仿佛走进一个洞穴，越陷越深，只看见光亮的洞口在远处越来越小。

你越来越害怕，向DAWN求救："卡丽丝是不是把我们的轨道算错了？我们是不是陷入黑洞视界了？我们要完了吗？！"

"提克哈依，提克哈依，"她安慰你，"我们没有危险，我们还在视界外面。黑暗笼罩整个天空，不过是黑洞引力的强烈透镜作用。看那儿，我指的地方，差不多就在头顶上，那是星系3C295。你下落之前，它还在水平的位置，离天顶90°。但是在这儿，'巨人'的视界附近，黑洞引力强烈作用在来自3C295的光线上，使它们从水平偏转到几乎垂直，结果，3C295就出现在我们头上。"




图P.4　飞船漂在黑洞视界的上方，光通过那些轨道从遥远星系来到视界。黑洞引力使光线向下偏转（"引力透镜效应"），飞船上的人看见所有的光都汇聚成头上的一个圆形亮点。

你放心了，继续下降。工作台显示了飞船经过的径向（向下）距

离和通过你们位置的绕黑洞的圆轨道的周长。刚开始时，每径向下落1千米，轨道周长减少6.283 185 307…千米，周长减少与半径减小的比为6.283 185 307千米：1千米，它等于2π，这正是欧几里得的标准圆周公式所预言的。但是现在你的飞船邻近视界，周长减小与半径减小的比比2π小得多：在10倍视界周长处，它是5.960 752 960；2倍处，是4.442 882 938；1.1倍处，是1.894 451 650；1.01倍处，是0.625 200 306。只有在弯曲空间里，才会出现与你在十几岁时学的标准欧几里得几何相差如此巨大的偏离；你现在看到的是爱因斯坦广义相对论所预言的与黑洞的潮汐力相伴的曲率。[3]

在最后阶段，卡丽丝需要费越来越大的力量才能靠火箭使飞船的降落速度慢下来，终于，飞船来到1.000 1个视界周长的轨道，凭着10g的向上加速度克服了黑洞强大的引力，静静地漂在视界的上方。它下落最后1公里时，周长只减小0.062 828 712千米。

船员们忍着10个地球重力的痛苦，拿出望远镜摄影机，投入周密的摄影工作。除了你们周围有一点儿因为下落气体碰撞生热而产生的微弱辐射外，要拍摄的电磁波都在头顶那个亮点里。那亮点很小，直径只有3弧度，是从地球看到的太阳大小的6倍。[10]但细看下去，那儿是围绕着“巨人”的所有恒星和宇宙中所有星系的像。出现在亮点正中心的星系是真正在头顶上的。从中心到边缘的55%，是像3C295那样的星系的像，假如没有黑洞的透镜效应，它们应该在水平位置，离天顶90°。从这里到边缘的35%，是在黑洞另一边，即在我们正下方的那些星系的像。最外面的30%，是每个星系的第二次像；而最外面的2%，是第3次像！同样奇怪的是，所有恒星和星系的颜色都是假的。你知道的某个星系本是绿色的，而现在它似乎闪烁着微弱的X射线：“巨人”的引力把这个星系的辐射引向你们，使它增大了能量，波长从$5\times10^{-7}$米（绿光）减到$5\times10^{-9}$米（X射线）。同样，类星体3C273的外缘，你知道原来发射波长为$5\times10^{-5}$米的红外辐射，现在看到它闪着波长为$5\times10^{-7}$米的绿光。[11]

完整记录了头上的亮点后，你们开始关心飞船的内部。你们几乎都以为，在这黑洞附近，物理学定律会有某些改变，而这些改变也会影响每个人的生理。情况并不如此。你看大副卡丽丝，她显得很正常；再看二副布里特，他也很正常。你们握握手，你也感觉正常。你喝一杯水，除了10g的效应外，也跟平常一样。卡丽丝打开氩离子激光器，跟过去一样，它发出明亮的绿光；布里特发出一束红色激光脉冲，测量它从激光器到镜子然后返回所用的时间，再根据测量计算光的速度，结果与地球实验是绝对一样的：每秒299 792千米。

船里的一切事情都正常，仿佛它就停在一个具有10g重力的大质量行星表面。假如不向外看飞船头上那个怪异的亮点和周围吞噬一切的黑暗，你不会知道——或几乎不会知道，你正在一个黑洞视界的邻近，而完全不是在某个行星的表面。飞船里的时空跟外面的一样也会被黑洞弯曲，通过足够精确的测量，你可以测出它的曲率，例如，你可以测量头脚之间的潮汐拉伸。但是，尽管时空曲率在视界300万亿千米周长的尺度上起着巨大作用，在你那1公里的飞船尺度上，它的效应却小得可怜。曲率在飞船两端产生的潮汐力只是地球引力的百万亿分之一（$10^{-14}$g），而你头脚间的力还要小1 000倍！

这种正常也是值得留意的。为了进一步认识它，布里特从飞船放出一只太空舱，为了测量光速，让它带着脉冲式的激光器和反射镜。太空舱落向视界时，仪器测量了光脉冲从舱头的激光器到舱尾的反射镜然后返回的速度。太空舱的计算机把计算结果通过激光束传向飞船："每秒299 792千米；299 792；299 792…"当太空舱离视界越来越近时，回来的激光的颜色也从绿移到红到红外到微波、无线电波……，但所载信号都是一样的："299 792；299 792；299 792…"然后，激光消失了。太空舱越过了视界，它里面的光速在它下落时也从来没有发生过改变，决定它那些电子系统运行的物理学定律也没有任何改变。

你对这些实验结果非常满意。在20世纪初，爱因斯坦曾宣告（他

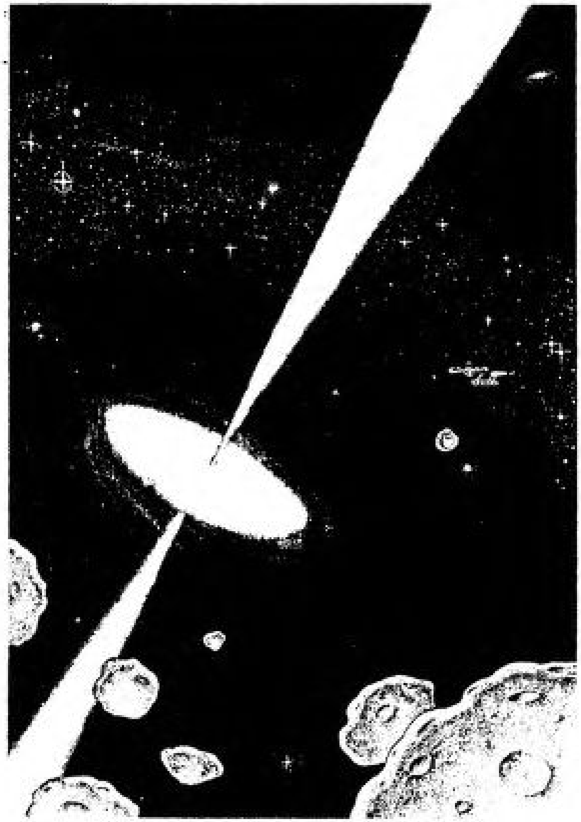
主要从哲学上考虑），局部的物理学定律（即定律所在区域很小，可以忽略时空曲率）在宇宙中应该是处处一样的。这个宣言被尊为物理学的一个基本原理：等效原理。[4]在后来的世纪里，等效原理常常经受实验的检验，但它还从来没有经历过像你们在“巨人”视界邻近做的实验那么生动而彻底的检验。

10个地球重力令你和你的船员们疲惫了。于是，你们准备航行的最后一步，回银河系。在航行之初，船员会把你们的“巨人”探险报告发回去；由于飞船很快也会近光速旅行，所以，从地球看来，报告会比飞船早一年到达银河系。

飞船升起离开“巨人”的时候，你的船员仔细用望远镜研究了头上的类星体3C273（图P.5）。[5]从类星体中心射出两股巨大的尖尖的热气体喷流，300万光年长。将望远镜瞄准中心，你们看到了喷流的源泉：一个厚厚的热气体环，大小不足1光年，黑洞在环的中心。这个被天体物理学家称为“吸积盘”的环一圈圈地绕着黑洞。船员们测量了它的旋转周期和周长，推测黑洞质量是20亿（$2\times10^9$）太阳质量，比“巨人”小7 500倍，但远远大于银河系里的任何黑洞。在黑洞引力作用下，气流从环流向视界；接近黑洞时，你们会看到以前不曾见过的现象：气流像龙卷风一样绕着黑洞盘旋——黑洞一定在快速旋转！旋转轴很容易确定：气流旋涡的轴就是黑洞旋转的轴。你发现，两股喷流是沿着转轴射出来的。它们就在视界的南北两极生成，从黑洞的旋转和气体环中汲取能量，[6]就像龙卷风从大地卷起尘埃。

你很奇怪，为什么“巨人”与3C273有那么大的不同：为什么质量和尺度都大1000倍的“巨人”没有环绕的气体圈和巨大的类星体喷流？布里特经过长时间的望远镜观测，找到了答案：每过几个月，就会有一颗在环绕3C273的小黑洞的轨道上的恒星坠向视界，被黑洞潮汐粉碎；恒星内约1个太阳质量的气体便喷射出来洒落在黑洞周围，在内摩擦力驱动下，慢慢进入气体环。这些新来的气体源源不断地补充着落进黑洞和喷流的气体。于是，气环和喷流总保持着丰富的气体来

源，能持续地发光。



图P.5　类星体3C273：气体环（“吸积盘”）包围的一个20亿太阳质量的黑洞，沿黑洞旋转轴射出两股巨大喷流。

恒星当然也会坠向“巨人”，布里特解释。但是，“巨人”远远大于3C273，它视界外的潮汐力太弱，不可能粉碎任何星体。恒星会完全被巨人吞没而不能喷出内部的气体形成环。因为没有气体环，巨人也就无法产生喷流和其他类星体的剧烈现象。

你的飞船继续上升，远离“巨人”的引力。你计划着回家的航行。回到银河系的地球时，距你们离开已经40亿年了。人类社会一定发生了巨大变化，你们不想回去了。你和船员决定在一个旋转黑洞的周围

开辟一块空间。你们知道，像3C273中的黑洞的旋转能可以为类星体喷流提供动力一样，一个小黑洞的旋转能也可以作为人类文明的能源。

你不想在某个黑洞看到已经有人在它周围建设了文明，所以，你的飞船没有飞向已经存在的快速旋转的黑洞，而是飞向某个恒星系统，在你到达不久，那儿会诞生新的快速旋转的黑洞。

你们离开地球时，银河系猎户座星云里有一个双星系，由两颗相互环绕的30个太阳质量的恒星构成。DAWN已经计算了，在你们去“巨人”时，那两颗恒星应该发生坍缩，分别形成一个24个太阳质量的无旋转黑洞（6个太阳质量的气体在坍缩中喷射出去了）。现在两个黑洞正相互环绕着，像一个双黑洞系；在环行中，它们会发出潮汐力的振荡（“时空曲率”的波动），也就是引力波。[7]像射出的子弹对枪有反冲作用一样，引力波也会对黑洞产生反冲，引力波反冲能减缓黑洞不可避免的螺旋下落的过程。你们稍稍调节一下飞船的加速度，就能赶上那螺旋下落的最后一幕：几天以后，你会看到两个黑洞无旋转的视界在绕着对方不停地旋转，越靠越近，越转越快，最后连在一起，形成一个更大的有旋涡的旋转视界。

原来的两个黑洞不旋转，不能作为你开拓的有效能源，不过，新生的这个快速旋转的黑洞却是很理想的！

[1]　第9章。

[2]　第8章。

[3]　第2，3章。

[4]　第2章。

[5]　第9章。

[6]　第9，11章。

[7] 第10章。

## 家园

经过42年的航行，飞船最后减速来到猎户座星云里DAWN预言那两个黑洞所在的地方。它们真在那儿。通过测量落向黑洞的星际原子的轨道运动，你证实了DAWN的预言，两个视界没有旋转，每个黑洞重24个太阳质量。每个视界的周长为440千米，相距30 000千米；黑洞每13秒绕对方转一圈。把这些数据代入广义相对论的引力波反冲公式，你认定两个黑洞将在7天后结合。[12]你的船员有充分的时间准备好望远镜摄像机，等着记录结合的细节。通过拍摄星光聚焦形成的黑洞盘外的光环，船员们很容易监测黑洞的运动。

你想走得更近，看得更清楚，而又能很安全地躲过黑洞的潮汐力。你决定，飞船最好落在比黑洞轨道大10倍的轨道上——一个直径300 000千米、周长940 000千米的轨道。卡丽丝把飞船引入那个轨道，船员们开始用望远镜进行摄影观测。

在接下来的6天里，两个黑洞越靠越近，轨道运动也越来越快。结合前一天，它们的距离从30 000千米收缩到18 000千米，轨道周期从13秒缩短到6.3秒。1小时前，距离是8 400千米，轨道周期是1.9秒。1分钟前，距离3 000千米，周期0.41秒。10秒前，距离1 900千米，周期0.21秒。

在最后10秒钟里，你和你的船开始摇晃了，先很轻，然后越来越剧烈。仿佛一双巨手在抓住你的头和脚，一会儿拉，一会儿压，劲儿越来越大，动作越来越快。不过，来得快，去得更快，一会儿就不摇了，一切又安静下来。

“怎么回事？”你向DAWN嘟哝，声音还在颤抖。

“提克哈依，提克哈依，”她安慰你说，“那是黑洞结合时产生的引

力波的起伏的潮汐力。你习惯了只有用精密仪器才能探测出潮汐力的弱引力波。而这儿是在结合的黑洞附近，引力波非常强大——我们飞船的轨道假如小30倍，它就会被波动粉碎。但我们现在很安全，黑洞结合完了，引力波飘走了，它们飞向宇宙，为遥远的天文学家带去黑洞结合的交响曲。”[1]

你把望远镜对着下面的引力源，看到真像DAWN说的，黑洞结合完了。过去有两个黑洞的地方现在只有一个。从下落原子的旋涡，你知道那个黑洞在快速地旋转，它将成为你的船员和他们千秋万代的子孙的理想发动机。

卡丽丝测量了飞船轨道，得出黑洞有45个太阳质量。原来的两个黑洞共48个太阳质量，那么一定有3个太阳质量转化成了纯能量，被引力波带走了。难怪那些波曾那样强烈地震撼着你！

当你调转望远镜对着黑洞时，一个意外的东西从船外飞过，光亮向四面散开，然后在你的船边炸开一个洞。训练有素的船员和机器人立即各就各位，准备战斗，却没有发现攻击你们的敌船——于是，你又请DAWN来帮忙。她通过飞船的语音系统安慰大家：“提克哈依，提克哈依。我们没有遭遇攻击。那不过是一个怪异的原生黑洞在蒸发，然后爆炸了。”[2]

“什么？！”你喊了起来。

“一个原生黑洞，蒸发了，然后在爆炸中毁灭了。”DAWN回答。

“说明白些！”你命令，“你说原生是什么意思？你说蒸发和爆炸是什么意思？你在说废话。东西可以掉进黑洞，但没有东西能出来；没有什么能‘蒸发’。黑洞会永远存在，它总在增大，永不收缩。黑洞不可能‘爆炸’，不可能毁灭自己。那太离奇了！”

DAWN还是那么有耐性，她告诉你，“大物体——如人、恒星和

恒星坍缩形成的黑洞——都是经典的物理学定律决定的，如牛顿运动定律、爱因斯坦广义相对论定律等。相反，小物体——如分子、原子和比原子还小的黑洞——是由一组大不相同的量子物理学定律决定的。[3]经典定律严禁正常大小的黑洞蒸发、收缩、爆炸和毁灭，但量子定律不像这样，它们要求任何原子大小的黑洞慢慢蒸发、收缩，直到某个原子核大小的临界周长。这样的黑洞虽然小，却重达几十亿吨，那时它必然会在巨大的爆炸中毁灭自己。几十亿吨的质量通过爆炸转化为向外喷发的能量，比20世纪人类在地球上爆炸的最大的核武器的能量还大1万亿倍。刚才损坏我们飞船的就是这样的爆炸。”

“不过你不必担心会有更多的爆炸，”DAWN接着说，“因为小黑洞极少，所以这样的爆炸也很罕见。小黑洞都是200亿年前在宇宙大爆炸中生成的，这就是为什么它们叫原生黑洞。大爆炸只产生了那么些原生黑洞，而那些原生黑洞自诞生以来一直在慢慢地蒸发。偶尔会有个别黑洞达到最小临界尺度而爆炸。[4]而一个黑洞在经过我们的船时爆炸，是极不可能发生的事情——我们不过碰巧遇到了；而且，我们的船极不可能再碰到这样的黑洞了。”

你感觉轻松了，命令船员开始修理，而你和助手们则开始用望远镜观测你们下面那个45个太阳质量的快速旋转的黑洞。

黑洞的旋转不仅表现在螺旋下落的原子，还表现在你们下面那个光环环绕的黑点的形状；那黑点像个扁南瓜，赤道隆起，两极平坦，正是黑洞旋转的离心力产生的结果。[5]但赤道隆起并不对称，盘的右边（黑洞旋转时离开你的那一边）显得比左边更大。据DAWN的解释，视界更容易捕获沿它右边向着你来而对着它旋转方向的星光，不太容易捕获从左边来的顺着它旋转方向的光。

布里特测量了黑点的形状，并与广义相对论的黑洞公式做了比较，发现黑洞旋转的角动量是它的质量所能允许的角动量的96%。根据这样的角动量和黑洞的45个太阳质量，你计算了其他一些性质，包

括它的旋转速率，每秒270周，和它的赤道周长，533公里。

你对黑洞的旋转很感兴趣，以前从来不可能这么近地观察旋转黑洞。所以，虽然很过意不去，你还是请一个志愿者机器人到视界近旁去探险，并把经过发回来。你对那个机器人（他叫科罗（Kolob））仔细下达了指令："降到视界上10米的地方，靠你的火箭使自己静止下来，浮在飞船正下方。还要靠你的火箭抵抗引力的向下吸引和空间的龙卷风旋涡。"

科罗喜欢冒险，他离开船舱，向下落去。起先，他轻轻点燃火箭，就能抵抗空间的旋涡，让自己保持在飞船下面，但后来就困难了。当他到达的轨道周长为833公里，比视界大56%时，他的激光带回了这样的消息："我顶不住旋涡；我顶不住了，顶不住了！"他像被龙卷风卷起的一块石头，被卷入了围绕黑洞的轨道。[6]

"别担心，"你告诉他，"尽可能顶住旋涡，继续降落，直到视界上方10厘米。"

科罗答应了。他接着下落，被卷入越来越快的环行运动。最后，他停止下落，漂在视界上方10米的地方，却几乎与视界本身同步地飞旋着，每秒钟270圈。不论费多大劲，都挡不住这种运动，因为空间旋涡，他永远也停不下来。

"换一个方向加速。"你命令。"既然不能比每秒270圈转得更慢，那你就转快一些。"

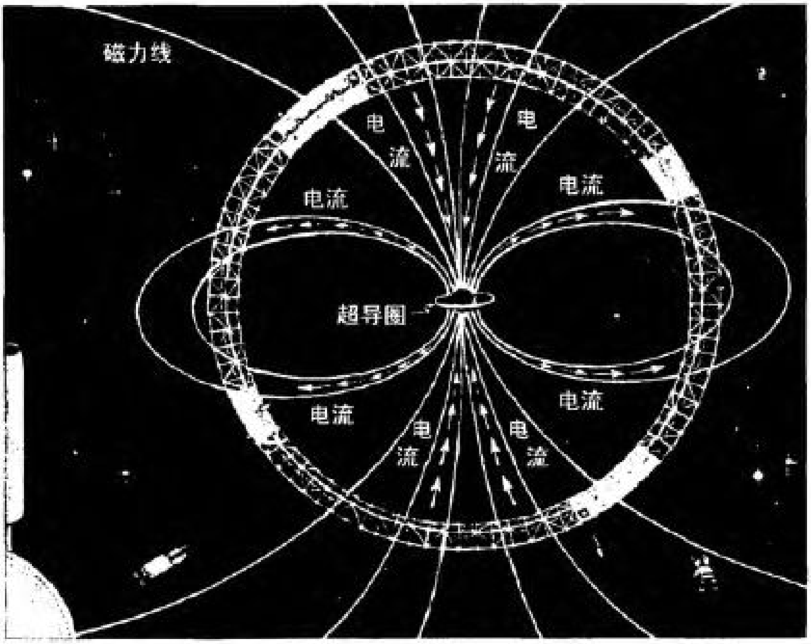
科罗试了试。他加速火箭，想让自己还在视界上方10米，但比先前运行更快。尽管他从火箭感觉到了平常的加速度，但你看他的运动却几乎没有什么改变。他仍然每秒钟环行270圈；在你还没来得及给他发出进一步指令时，他的燃料用完了，开始垂直下落；他发出的激光突然地掠过电磁波谱，从绿变红，到红外，到无线电波，然后变

黑，而他的飞行却没有改变。他去了，落进了黑洞，落向你永远也看不到的暴戾的奇点。

经过3个星期的痛苦、实验和望远镜观测，你们现在开始建设未来了。从遥远的行星取来材料，在黑洞周围建起环状“大梁工程”，周长500万千米，厚3.4千米，宽4　000千米。它旋转的速度恰到好处，每小时转两圈，这样，离心力正好能抵消大梁环中心（距里外两面各1.7公里）受到的黑洞引力。环的大小也是仔细考虑过的，喜欢1个地球重力的人可以在环的里面和外面建设家园，喜欢重力轻一点的人可以住在中心附近。引力的差别，部分来自旋转环的离心力，部分来自黑洞的潮汐力——用爱因斯坦的话说，即时空曲率。[13]

为这个环状世界提供光和热的电源来自黑洞：黑洞质量的20%以能量形式贮藏在视界附近的空间的龙卷风式的旋涡里。[7]那是太阳一生所辐射的光和热的10　000倍！因为在视界外面，那是能够提取的。即使环状世界只能利用50%的能量，也仍然比太阳的能量供应大5　000倍。

能量的汲取原理与类星体是相同的：[8]船员们将磁场穿过黑洞视界，虽然它有离开的趋向，你们还是利用巨大的超导感应圈（图P.6）将它留在黑洞。视界旋转时，在附近的空间产生龙卷风旋涡，它反过来又与穿过的磁场相互作用而形成巨大的发电机。磁力线充当着输电线。电流从黑洞赤道流出（表现为电子从这里流进），沿着磁力线流向环状世界，将能量送到那儿。然后，它沿着别的磁力线离开环状世界，从南北两极流进黑洞（表现为质子从那儿流进）。通过调节磁场强度，环状世界的居民可以调节能量输出：早期的磁场弱，能量小；晚期的磁场强，能量大。随着能量的汲取，黑洞旋转会逐渐变慢，但仍然要过亿万年它才能耗尽所贮藏的巨大旋转能。




图P.6　围着黑洞的大梁环上的城市和城市从黑洞的旋转汲取能量的电磁系统。

这个人造的世界就是船员的“家园”，是他们子孙万代的家园，也是他们未来探索宇宙的基地。但是，你不喜欢这儿，你怀念地球和地球上的朋友，他们一定已经死去40多亿年了。你真想在你200年生命的最后1/4回到如诗如画的青年时代，那是很冒险的，也许不会有结果，但你还是想试试。

走向未来是很容易的，如你们经历的黑洞航行；回到过去却没那么简单。实际上，物理学的基本定律也许完全禁止这样的旅行。不过，DAWN告诉你，20世纪的物理学家曾猜想，通过一种叫虫洞的假想的空间卷曲，也许可以实现回到过去的时间旅行。[9]这种空间卷曲由两个人口（虫洞口）构成，像两个没有视界的黑洞，在宇宙中可以分离很远（图P.7）。从一个洞口进去的东西会发现一个很短的通道（虫洞的喉），通向另一个洞口。这条通道在超空间延伸，不穿过正常空间，所以从我们的宇宙看不到它。DAWN解释，通过虫洞的时间

与通过我们宇宙的时间，在连结方式上可能大不相同。沿一个方向穿越虫洞，如从左到右，人们可能回到宇宙的过去，而从反方向穿越，即从右向左，他可能会跑到时间前头。这样的虫洞不仅是空间卷曲，也是时间卷曲的结果。



图P.7　一个假想虫洞的两个洞口。从一个洞口进去，穿过一条短短的（虫洞喉）在超空间而不是我们宇宙中的通道，你会从另一个洞口出来。

DAWN告诉你，量子引力定律要求，应该存在这种类型的非常微小的虫洞。[10]这些量子虫洞的大小只有$10^{-33}$厘米，它们的存在也只是瞬间的事情——短短的$10^{-43}$秒，当然不能用来作时间旅行。[14]它们出人意料地闪现，又出人意料地消失——忽来忽去，又似乎无处不在。碰巧，可能有个虫洞，一个洞口在今天的环状世界附近，另一个洞口在40亿年前你们启程远航时的地球附近。DAWN建议，在虫洞闪现时抓住它，然后像小时候吹气球那样让它膨胀，保持洞口打开，让你穿过它回到年轻时的故乡。

但DAWN也警告你，那是很危险的。物理学家猜想（尽管还没有证明），在膨胀的虫洞成为时间机器前的那一瞬间，它可能就在剧烈的爆炸中自我毁灭了。宇宙可能通过这样的办法来阻止它自己出现时

间旅行的怪圈，例如，一个人可以回到过去，在母亲怀他之前将母亲杀死，从而不让他出生来杀害母亲。[11]

如果物理学家猜错了，DAWN就可以让虫洞打开几秒钟并张开足够你穿过的喉管。你在旁边等着，然后钻进去，经过几分之一秒（你自己的时间），你就回到了40亿年前你年轻时在地球的家乡。但是，假如时间机器自我毁灭了，你也会随它而去。你决定碰碰运气……

上面的故事像科幻小说，是的，的确有点儿像。我无法保证织女星旁有10个太阳质量的黑洞，银河系中心有100万个太阳质量的黑洞，或者宇宙什么地方有15万亿个太阳质量的黑洞；这些都是虚构的，然而却是合理的。我自己也怀疑，人类是否有力量成功进行星系际旅行，或者星际旅行，他们是否能在黑洞的周围建成大梁上的环状世界。这些也是虚构的。

不过，我能很有信心（当然还不能彻底）地保证，我们的宇宙存在着黑洞，它们具有故事里描述的那些性质。假如你的飞船漂浮在15万亿个太阳质量的黑洞视界上方，我保证船里的物理学定律与地球上的是一样的；当你看船外周围的天空时，你会发现整个宇宙都暗下来了，只有一个明亮的小光盘在照着你。我保证，假如你让一个机器人到旋转黑洞的附近去探险，不论它如何发动火箭，都只能以黑洞自身的旋转速度（在我说的例子中，即每秒270周）进退。我保证，快速旋转的黑洞能将它质量的29%作为旋转能贮藏起来，如果我们足够聪明，是能汲取和利用它的。

我从没见过黑洞，怎么能有信心保证这些事情呢？实际上，没人见过黑洞，天文学家也只发现了一点儿间接的黑洞存在的证据，[12]而关于它们的那些具体性质，什么观测证据也没有。我凭什么那么大胆地保证那么多的东西呢？原因很简单。假如我们理解正确的话，物理学定律预言那些黑洞性质，而且是毫不含糊地预言，实际上跟它们预言地球上的海洋潮汐（每次高潮和低潮的时间和高度）是一样的。根

据牛顿的物理学定律，可以从数学公式导出从1999年到2010年的地球潮汐序列；同样，根据爱因斯坦的广义相对论定律，可以从数学计算导出黑洞视界和外面的一切性质。

我为什么相信物理学基本定律的广义相对论描述是高度精确的呢？毕竟，我们知道牛顿的描述在黑洞附近不再准确了。

基本定律的成功描述本身都暗示着它会在什么地方失效。[13]牛顿的描述告诉我们，它可能在黑洞附近失效（当然，我们只是在20世纪才从牛顿的描述中发现这一点）。同样，爱因斯坦的广义相对论描述的可靠性表现在黑洞外、视界上和几乎一切（但不完全）都落向它的中心奇点的黑洞内部。这是令我相信广义相对论预言的一个方面；另一方面的事实是，虽然广义相对论的黑洞预言还没有被直接检验过，但广义相对论的其他特征已经在地球上、在太阳系、在由两颗致密奇异的所谓脉冲星构成的双星系中找到了高度精确的验证。广义相对论成功经历了每一个考验。[15]

在过去的20年里，我参与了有关的理论物理学探索，得到了现在这些黑洞的认识，我也在探索通过天文学观测来检验黑洞的预言。我个人的成绩是渺小的，但与物理学家和天文学家同行在一起，我经历了探索的兴奋和发现的惊奇。我想尽可能地在这本书里把那些兴奋和惊奇的感觉带给天文学家和物理学家以外的朋友们。

[1]　第10章。

[2]　第12章。

[3]　第4～6，10，12～14章。

[4]　第12章。

[5]　第7章。

[6]　第7章。

[7]　第7，11章。

[8]　第9，11章。

[9]　第14章。

[10]　第14章。

[11]　第14章。

[12]　第8，9章。

[13]　第13，14章。

# 第1章　空间和时间的相对论

爱因斯坦推翻了

牛顿的绝对空间和绝对时间的概念

1901年4月13日

德国·莱比锡

莱比锡大学

威廉·奥斯特瓦尔德教授

尊敬的教授先生！

请您原谅一个父亲，为了儿子的事情冒昧地来打扰您，尊敬的教授先生。

我先应该告诉您，我的儿子阿尔伯特今年22岁，他在苏黎世综合技术学校（Zurich Polytechnikum）学了4年，去年夏天，他成功地通过了数学和物理学的学位考试。从那以后，他一直想找一份助教的工作，这能帮助他继续学习理论和实验物理学，但是，还没有找到。在他求职的时候，人们给他的评语都称赞他的才干；无论如何，我可以向您保证，他非常好学，非常用功，非常爱他的科学。

所以，我儿子为现在还没有工作而深感不幸，而且他一天天地确信，他已经脱离了事业的轨道，而现在又没人同他来往。另外，他还认为他是我们的负担，一个没用的人，这使他感到压抑。

您，尊敬的教授先生，在目前活跃在物理学界的学者中，是我儿子最崇拜和尊敬的一个；所以，我冒昧地请您满足我的一个小小的要求，看看他发表在《物理学纪事》上的论文，如果可能，请您为他写几句鼓励的话，使他重新在生活和工作中快乐起来。

另外，如果您现在或者今年秋天能为他找一个助教的职位，我将无限感激。

我再次请您原谅我唐突地给您写信，我还要冒昧地说一句，我儿子一点儿也不知道我的无礼举动。

就说这些，尊敬的教授先生。您忠实的

赫尔曼·爱因斯坦[1]

是的，有一段时期，阿尔伯特·爱因斯坦真的很沮丧。自21岁从苏黎世综合技术学校毕业以来，8个月没有工作，他觉得自己失败了。

在综合技术学校（Polytechnikum，通常根据德文的第一个字母称为“ETH”[1]），爱因斯坦曾跟世界最著名的几个物理学家和数学家学习，但同他们的关系并不融洽。在世纪之交的科学世界里，大多数的教授（Professor）都要求和希望别人尊重他们，而爱因斯坦却并不那么尊重他们。因为还在小时候他就敢反权威，没经他自己亲自检验过的任何事情，他从不接受，而总要提出问题。他断言，“对权威的盲从是真理的最大敌人。”[2]他在ETH的两个最著名的物理学教授之一的韦伯（Heinrich Weber）曾恼火地抱怨爱因斯坦，你是个聪明的孩子，非常聪明的孩子。但你有一个最大的缺点：“你听不进别人的任何东西。”他的另一个物理学教授佩内特（Jean Pernet）问他为什么不学医学、法律或者哲学，而偏学物理，“你可以做你喜欢的事情，”佩内特说，“我警告你都是为了你好。”

由于爱因斯坦对功课漠不关心，情况并没有好起来。“为了考

试，不管你是不是喜欢，都得把所有的东西塞进脑子里。”他后来这么说。他的数学教授闵可夫斯基（Hermann Minkowski，在第2章里我们会更多地听到这个名字）对爱因斯坦的这种态度非常生气，说他是一只“懒狗”。

然而，爱因斯坦并不懒，他不过是有自己的选择。有些功课他全部都吸收了，而另外一些他忽略了。他更喜欢把时间花在自学和独立思考上。思考是一种乐趣，令人愉快，能带来满足。他可以靠自己学习“新”物理学，而这些物理学在韦伯的所有课程中却被省略了。

## 牛顿的绝对空间和时间，以太

“旧”物理学，即爱因斯坦能够从韦伯那儿学到的物理学，是一个庞大的知识体，我称它是牛顿的物理学，这并不因为它完全属于牛顿（他做不了这么多），而是因为它的基础是牛顿在17世纪奠定的。

19世纪后期，物理宇宙间的一切迥然不同的现象，都可以通过几个简单的牛顿物理学定律得到优美的解释。例如，所有与引力有关的现象都可以用牛顿运动和引力定律来解释：

·不受力作用的物体将沿直线匀速运动。

·在力的作用下，物体速度发生改变，变化率与力成正比，与物体质量成反比。

·宇宙间任意两个物体间存在着引力作用，它与物体质量的乘积成正比，与物体间的距离的平方成反比。

运用这三个定律的数学操作，[2]19世纪的物理学家可以解释行星绕太阳的轨道，卫星绕行星的轨道，海洋的潮汐和岩石的崩落，他们甚至知道怎么去称太阳和地球的质量。同样，运用一组简单的关于电

和磁的定律，物理学家们可以解释闪电、磁铁、无线电波以及光的传播、衍射和反射。

名声和财富在等着那些在技术上运用牛顿定律的人。瓦特（James Watt）通过牛顿热定律的数学运算，提出如何将别人设计的原始蒸汽机改造为实用的机器，那就是后来以他名字命名的蒸汽机。莫尔斯（Samuel Morse）靠亨利（Joseph Henry）的帮助理解了电磁定律，发明了他的有很高实用价值的电报码。

物理学家跟发明家一样，都为他们能完美地理解宇宙而感到自豪。天地间万物似乎都遵从牛顿的物理学定律。人类征服了这些定律，它们也正引导人类去征服他们的环境——也许某一天，还会去征服整个宇宙。

所有这些旧的牢固确立的牛顿定律及其技术应用，爱因斯坦在韦伯的课中都学到了，而且学得很好。实际上，在ETH的最初几年，爱因斯坦是很欢迎韦伯的。1898年2月，他在给班里惟一的女生米列娃（Mileva Maric'，他爱上她了）的信中写道："韦伯的课讲得精彩极了，我急切地盼着听他的每一堂课。"[4]

但是到ETH的第四年，爱因斯坦不满意了。韦伯只讲了旧物理学，他完全忽略了近几十年来一些最重要的物理学进展，连麦克斯韦（James Clerk Maxwell）新发现的一组精妙的电磁学定律也忽略了。从麦克斯韦的定律，人们可以导出所有的电磁现象：磁体的行为、电火花、电流、无线电波、光。爱因斯坦只好通过阅读其他大学的教授写的最新著作来自学麦克斯韦的统一的电磁学定律，他大概还直率地向韦伯表示了他的不满，两人的关系恶化了。

追溯起来，韦伯在他的课程里忽略了太多的东西，而其中最重要的显然是他忽略了，越来越多的证据表明，牛顿物理学大厦的基础出现了裂缝，而这个基础的砖块和砂浆，就是牛顿的绝对空间和绝对时

间的概念。

牛顿的绝对空间是日常经验的空间，它有三维：东-西、南-北、上-下。日常经验告诉我们，有而且只有这么一个空间。它是全人类、太阳、所有行星和恒星所共同拥有的空间。我们都在这个空间里以自己的方式和速度运动，不论如何运动，我们感受空间的方式是一样的。这个空间让我们感觉长、宽、高，而依照牛顿的观点，不论如何运动，只要测量足够精确，我们对同一物体会得到相同的长、宽、高。

牛顿的绝对时间是日常经验的时间，时间像我们的岁月一样无情地流逝，我们用高质量的钟表，或者根据地球的转动和行星的运行来测量时间。全人类、太阳、所有行星和恒星，都共同经历着时间的流逝。依照牛顿的观点，不论如何运动，关于某个行星轨道的周期，或者某个政治家演说的时间，我们会得到一致的结果，只要我们都用足够精确的钟来测量。

如果牛顿的绝对空间和绝对时间的概念崩溃了，牛顿物理学定律的整个大厦就会倾覆。幸运的是，几年过去了，几十年过去了，两百年过去了，牛顿的概念基础依然牢固地屹立着。从行星天地到电的王国到热的世界，它赢得了一个又一个科学胜利。这个基础没有露出丝毫破裂的迹象——不过，到了1881年，情况不同了。这一年，迈克尔逊（Albert Michelson）开始测量光的传播。

如果我们测量光（或者别的什么东西）的速度，那么显然，测量结果似乎一定会依赖于我们的运动方式，而牛顿定律也是这么要求的。如果我们在绝对空间中静止，那么我们会看到光在各个方向上的速度是一样的。反过来，如果我们在绝对空间中运动，比如说向东运动，那么我们将发现向东传播的光会慢下来而向西传播的光会快起来，正如人们在向东行驶的列车上看到的那样，东飞的鸟慢了而西飞的鸟却快了。

对鸟来说，决定它们飞行速度的是空气。鸟在空气中扇动翅膀，不管朝哪个方向，它们都以相同的最大速度飞行。类似地，根据牛顿的物理学定律，决定光的传播速度的是一种被称为以太的物质。光在以太中振荡它的电场和磁场，不论沿什么方向，它总是以一个普适的速度在以太中传播。由于（照牛顿的观点）以太在绝对空间中是静止的，所以任何静止的人在所有方向上将测得相同的光速，而运动者会测得不同的光速。[5]

现在来看地球，它在绝对空间中穿行。不管别的，我们只考虑它环绕太阳的运动。1月，它沿某个方向运动，6个月以后（7月），它又运动到相反的方向。对应于这种运动，我们在地球上应该测到不同方向的不同光速，它们的差异应随季节而变化——尽管这个变化很小（大约只有万分之一），因为相对于光来说，地球的速度太慢了。

对实验物理学家来说，验证这一预言是一个很有吸引力的挑战。1881年，28岁的美国青年阿尔伯特·迈克尔逊用他自己发明的灵巧而精确的实验技术（现在叫“迈克尔逊干涉度量法”[3]）迎接了挑战。[6]迈克尔逊尽了最大努力，也没能发现任何有关光速随方向变化的证据。他在1881年的初次实验证明了光速在所有方向和任何季节都是相同的。1887年，迈克尔逊与化学家莫雷（Edward Morley）合作，在俄亥俄克里夫兰又进行了实验，以更高的精度证实了同样的结果。迈克尔逊很矛盾，既为他的发现高兴，也为结果感到失望。韦伯同19世纪90年代的其他大多数物理学家一样，对结果表示怀疑。

实验是很容易受怀疑的。有意义的实验常常是非常困难的——实在太难了，不论实验做得多么仔细，它们都可能产生错误的结果。哪怕是仪器的一点儿异常，或者温度的一点儿不可控制的波动，甚至仪器下地板的一点儿意外的振动，都会改变最后的实验结果。所以，一点儿也不奇怪，今天的物理学家同19世纪90年代的物理学家一样，偶尔也会碰到一些令人困惑的实验，这些实验要么互相矛盾，要么同我们对宇宙的本性和物理学定律的根深蒂固的信仰相矛盾。最近的例子

是，一些实验宣布发现了“第五种力”（在标准的高度成功的物理学定律中还没有出现过），而另一些实验却否定这种力的存在；还有实验宣布发现了“冷聚变”（这是标准的物理学定律所禁戒的现象，如果物理学家对这些定律理解正确的话），而又有实验否定冷聚变的发生。几乎所有威胁我们信仰的实验都是错误的，它们的基本结果都是实验误差的假象。不过，它们偶尔也可能是正确的，将为我们指明一条通向认识自然的革命道路。

杰出物理学家的一个标志是，他有能力“闻出”哪些实验可信，哪些实验不可信；哪些值得忧虑，哪些可以忽略。随着技术的改进和实验的多次重复，真理最终总会澄清；但是，如果谁想为科学进步作出贡献，想靠自己去确认那些重大发现，那么他必须预先而不是事后凭直觉判断哪些实验是可信的。

19世纪90年代的几位大物理学家审查了迈克尔逊-莫雷实验，他们认为，细致的实验装备和精心的实验操作，保证了实验是令人信服的。他们认定，这个实验“味道很好”，有理由认为牛顿物理学的基础出了问题。相反，韦伯和其他大多数人却相信，只要有时间，再进一步做些实验，一切都会明白，牛顿物理学将跟以往多次的经历一样，最终还会胜利。他们认为，即使在大学课程里提及这个实验也是不妥的，不能误导年轻人的思想。[7]

爱尔兰物理学家菲兹杰拉德（George F.Fitzgerald）第一个根据它的表面价值接受了迈克尔逊-莫雷实验，并考虑了它的意义。他拿这个实验同其他实验对比，[8]得到一个根本的结论：问题在于物理学家对“长度”概念的理解，相应地，牛顿的绝对空间的概念可能也存在错误。1889年，他在美国《科学》杂志的一篇短文中写道：

我以极大的兴趣阅读了迈克尔逊先生和莫雷先生奇妙而精巧的实验……他们的结果似乎同其他实验相矛盾……我想提出一个大概是唯一能够协调这种矛盾的假说，那就是，物体在通过以太[通过绝对空

间]或穿越它时，长度会发生变化，变化的量依赖于物体速度与光速之比的平方。

在沿地球运动的方向上，长度只发生了微小的收缩（十亿分之五），这可以（也确实能够）解释迈克尔逊-莫雷实验的零结果。[9]不过，这要求我们抛弃物理学家对事物行为的认识：没有什么已知的力能使运动物体在它们的运动方向上发生收缩，尽管收缩是那么微小。如果物理学家对空间的本性和对固体内部的分子力的认识是正确的，那么匀速运动的固体总会保持它在绝对空间中的形状和大小，而不管运动有多快。

阿姆斯特丹的洛伦兹（Hendrik　Lorentz）也相信迈克尔逊-莫雷实验，而且他特别重视菲兹杰拉德关于运动物体收缩的建议。菲兹杰拉德听说后，给洛伦兹写了封信，表示很高兴，“我因为自己的观点在这儿被嘲笑惨了。”为了更深入地理解，洛伦兹——还有法国巴黎的庞卡莱（Henri Poincaré）、英国剑桥的拉莫（Joseph Larmor），他们各自独立地重新考察了电磁学定律，发现了与菲兹杰拉德的收缩思想相吻合的一个特征：

如果我们以在绝对空间中静止的电场和磁场来表述麦克斯韦的电磁学定律，定律将具有特别简单而优美的数学形式。例如，一个定律的大意说在绝对空间静止的任何人看来，磁力线没有端点（见图1.1（a），（b））。然而，如果用一个运动者测量的稍微有点儿不同的场来表述麦克斯韦的定律，这些定律就会复杂而丑陋多了。特别是，“没有端点”的定律会变成，“在某些运动者看来，多数磁力线没有端点，但有些线被运动切断了，因而出现了端点。另外，当运动者挥动磁体时，新的磁力线又将被切断，然后联通，再切断，再联通。”（见图1-1（c））[10]

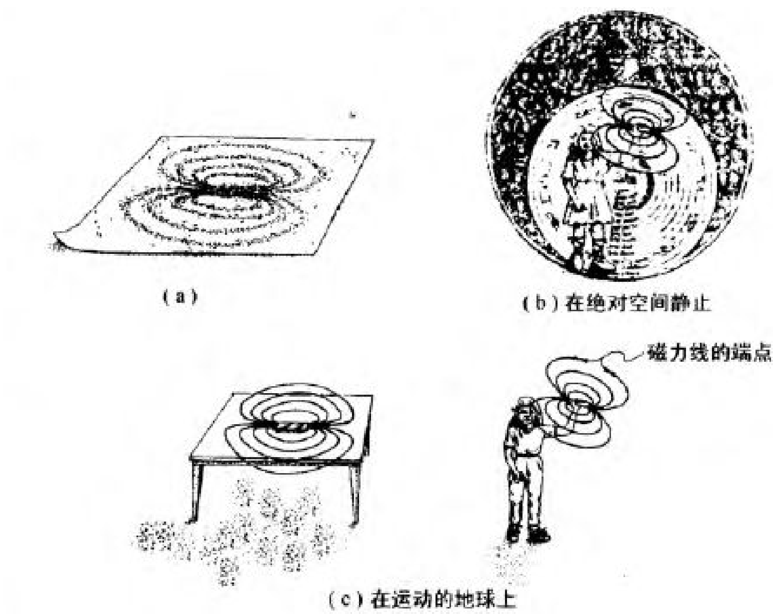



图1.1　在19世纪的物理学（即牛顿物理学）框架内认识的一个麦克斯韦电磁学定律：（a）磁力线的概念：在一张纸下放一块条形磁铁，在纸上洒些铁粉，则铁粉将显出磁力线。每根磁力线从磁铁的北极出发，绕着磁铁然后进入它的南极，穿过磁铁又到达北极，在北极自相联结。因此，磁力线是封闭曲线，没有端点，像橡皮圈。“磁力线永远没有端点”的说法，是麦克斯韦定律最简单、最漂亮的形式。（b）根据牛顿物理学，不论我们对磁铁做什么（例如，我们甚至可以大幅度挥动它），只要我们在绝对空间中静止，麦克斯韦定律的那种形式都是正确的。在静止观察者看来，没有磁力线会有端点。（c）根据牛顿物理学，以在绝对空间中穿行的地球上的人的认识来看，麦克斯韦定律要复杂得多。如果运动者的磁铁静止在桌面上，那么有一些力线（大约亿分之一）会出现端点。如果大幅度地挥动磁铁，则因挥动会有另外的力线（万亿分之一）被暂时切断，然后又联结，再切断，再联结。尽管任何19世纪的物理学实验都不可能辨别出小小的亿分之一或万亿分之一的有端点的力线，但在洛伦兹、彭加勒和拉莫看来，麦克斯韦定律产生这样的预言，本身就是复杂而丑陋的。

洛伦兹、庞卡莱和拉莫的数学新发现使运动者的电磁学定律也漂亮起来了，实际上，它们看起来跟在绝对空间中静止的人所用的定律是完全一样的：“不论在什么条件下，磁力线永远没有端点。”为了让定律都有这么漂亮的形式，我们只需要假定（与牛顿的戒律相反），所有运动物体在运动方向上发生收缩，而收缩的量正好精确地等于菲兹杰拉德为解释迈克尔逊-莫雷实验所要求的那个量！

如果说，菲兹杰拉德收缩只是我们用来让电磁学定律变得普遍地简单和优美的“新物理学”，那么洛伦兹、庞卡莱和拉莫呢？他们凭直觉相信，物理学定律本应是优美的，他们似乎已经抛弃了牛顿的戒律而坚定地相信收缩了。然而，只有收缩本身还不够，为使定律漂亮起来，我们还得假定，在宇宙中运动的人所测量的时间流比静止的人所测量的流更慢，运动让时间“膨胀”了。[11]

而在那个年代，牛顿的物理学定律是不容争议的：时间是绝对的。不论我们如何运动，时间总是以一个普适的速度无情地均匀地流逝着。如果牛顿定律是正确的，运动就不能使时间发生任何膨胀，正如它不能引起长度的任何收缩一样。不幸的是，19世纪90年代的钟远没有揭示这个事实的精度；另外，面对牛顿物理学在科学和技术上的胜利，而这些胜利又加固了绝对时间的基础，没有人愿意相信时间真会膨胀，洛伦兹、庞卡莱和拉莫不过是在空谈。

爱因斯坦这时还是苏黎世的一个学生，还没有准备好去解决这些令人兴奋的问题，不过他已经开始思考了。1899年，他给朋友米列娃（他对她的浪漫感情正在萌芽）写信说：“我越来越相信，今天这样的运动物体的电动力学是不对的。”[12]在接下来的6年里，随着物理学家能力的成熟，他将考虑长度收缩和时间膨胀的观点及其实在意义。[13]

相反，韦伯对这类思辨的想法一点儿也不感兴趣。他依然堂皇地讲他的牛顿物理学，似乎一切还是那么完美有序，似乎没有出现什么物理学基础的裂缝。

爱因斯坦在ETH的学习快要结束了，他很聪明，各科成绩也不是真的很坏（满分为6分，他的平均分是4.91），所以他天真地认为，他可以在韦伯手下当一名ETH的物理学“助教”，并像通常那样以此为跳板进入学术圈。如果做助教，他可以开始自己的研究，几年后获博士学位。

但结果并不是这样的。在1900年8月通过综合物理-数学科目最后考试的四个学生中，有三个得到了ETH数学家的助教职位，爱因斯坦是第四个，什么也没得到。韦伯请了两名学工程的学生做助教，没要爱因斯坦。

爱因斯坦继续想办法。毕业一个月后，在9月，他申请ETH的一个空缺的数学助教职位，被拒绝了。冬天和春天，他向德国莱比锡的奥斯特瓦尔德（Wilhelm Ostwald）和荷兰莱顿的昂内斯（Heike Kamerlingh Onnes）写过申请，却似乎连礼节性的回信也没有收到过——尽管，他给昂内斯的信现在骄傲地陈列在莱顿的博物馆里，而奥斯特瓦尔德在10年后会第一个提名爱因斯坦获诺贝尔奖。甚至爱因斯坦父亲给奥斯特瓦尔德的信似乎也没有回音。

米列娃活泼漂亮，意志坚强，爱因斯坦对她的感情更强烈了。[14]1901年3月27日，他在给她的信中说，“我绝对相信，事情都怪韦伯……给别的教授写信一点儿用都没有，因为他们一定会向韦伯打听我的某些事情，而他只会说我的坏话。”[15]1901年4月14日，他写信给亲密伙伴格罗斯曼（Marcel Grossmann），“如果不是韦伯在背后玩儿花样，我早就可以找到[助教职位]了。尽管如此，我会尽力的，也不会放弃我的幽默……上帝创造了蠢驴，还给他一身厚皮。”[16]

他真需要一身厚皮，这不仅是因为他没找到工作，还因为他父母强烈反对他同米列娃结婚，而他跟米列娃的关系也正面临着风暴。关于米列娃，他母亲说，“玛利奇小姐给我带来了我一生最痛苦的时刻，如果我能作主，我会尽一切力量让她从我们的眼前消失，我实在不喜欢她。”[17]而米列娃说爱因斯坦的母亲，“那老太太不仅想尽办法让我的生活痛苦，也让她儿子痛苦，这似乎就是她为自己设计的生活目标……我没想到，竟会有这种没心没肺的恶人，真是坏透了！”[18]

爱因斯坦绝望了，他想摆脱对父母的经济依赖，想有平和的心境和自由，好将更主要的精力投到物理学中去。也许他能通过别的途径

实现这个愿望，而不一定靠在大学当助教。他的ETH学历使他有资格在预科学校（高中）教书，所以他这么做了：1901年5月中旬，他设法在瑞士温特图尔的一所高等技术学校找了一份临时工作，代一位要服兵役的老师教数学。

爱因斯坦在给他的ETH历史老师斯特恩（Alfred Stem）的信中写道："我[因为教书的工作]高兴得快发狂了，因为我今天接到消息说一切都安排妥了。至于谁那么好心把我推荐到那儿去，我一点儿也不知道，因为有人告诉我，我从没上过以前任何一个老师的荣誉簿。"[19]继在温特图尔后，1901年秋他又临时在瑞士沙夫豪森的一个高中教书，然后，1902年6月，他成为瑞士专利局的一名"三级技术员"，从而独立了，也稳定了。

尽管爱因斯坦在个人生活上接连遭遇风波（他长期与米列娃分离；1902年同米列娃生了一个女儿，也许是为了让爱因斯坦能在保守的瑞士保住工作，他们将孩子当养子抚养；[20]一年后，他不顾父母的强烈反对，跟米列娃结婚了），他仍然保持着最佳的精神状态和足够清醒的头脑去思考物理学问题：从1901年到1904年，通过对在液体（如水）和在金属中分子之间的力的研究及对热的本性的研究，他锻炼了自己作为物理学家的技能。他那些新颖而且基本的发现，通过5篇论文相继发表于20世纪初最权威的物理学杂志：《物理学纪事》（Annalen der Physik）。

在伯尔尼专利局的工作很好地培养了爱因斯坦的才能。在专利工作中，他得向别人指出那些提交上来的发明是否有意义——这通常是令人愉快的事情，而这些工作也使他的思想变得敏锐起来。工作之余，还有一半的自由时间和整个周末，他大部分都用来学习和思考物理学了，[21]而且还经常处在家庭的喧嚣中。

不论多大干扰，他总能集中精力。一个在他同米列娃结婚几年后去过他家的学生描述了他的这种能力："在书房里，他坐在一堆满是

数学公式的稿子前面，右手写字，左手抱着小儿子，还不断回答正在玩儿积木的大儿子阿尔伯特提的问题。‘等会儿，马上就完了，’说着，他把孩子交给我看几分钟，又继续工作了。”[22]

爱因斯坦在伯尔尼与其他物理学家没有往来（不过他确实有几个很亲密的不是物理学家的朋友，他可以同他们讨论科学和哲学）。对大多数物理学家来说，孤立是一种灾难，他们需要不断与在相同问题上进行研究的同事联系，以免自己的研究会因迷失方向而徒劳无获。但爱因斯坦的智力与众不同，他在孤独中获得的成果比在其他物理学家激发的环境下更多。

有时，同别人的交谈对他也有帮助——那不是因为他们为他带来了什么新颖深刻的见解或信息，而是因为他通过向别人解释疑难和问题，可以在自己头脑中澄清这些疑问。对他帮助特别大的是贝索（Michele Angelo Besso），一个意大利工程师，曾经是他在ETH的同学，而现在同他一起，也在专利局工作。关于贝索，爱因斯坦说：“在整个欧洲，我再也找不到更好的知音了。”[23]





左：爱因斯坦坐在瑞士伯尔尼专利局的办公桌旁（约1905年）；右：爱因斯坦与妻子米列娃和儿子汉斯·阿尔伯特（约1904年）。[左，耶鲁撒冷希伯来大学爱因斯坦档案馆提供；右，伯尔尼瑞士联邦爱因斯坦学会文献/档案馆提供。]

[1]　ETH，Eidgenoessiche Technische Hochschule，照此，学校应称为“〔苏黎世〕联邦理工大学”。——译者

[2]　想知道“数学操作”这些物理学定律是什么意思的读者，可以在书后的注释部分找到有关讨论。[3]

[3]　见第10章。

# 爱因斯坦的相对空间和时间，绝对光速

对爱因斯坦来说，贝索在1905年5月给他的帮助是特别有意义的。那时，爱因斯坦从用心了几年的其他物理学问题回到麦克斯韦的电动力学定律和那些关于长度收缩和时间膨胀的诱人线索上来。他想找一个办法来为这些线索赋予意义，但思想遇到了障碍。为了清除绊脚石，他请贝索帮忙来了。据他后来回忆，“那是一个明媚的日子，我去找[贝索]，开门见山地对他说：‘我最近遇到一个问题，太难了，我理解不了，所以我今天带着问题到这儿来，跟你讨论。’我同他谈了很多，后来我突然明白是怎么回事了。第二天我又去找他，开口就说：‘谢谢你，我已经完全解决这个问题了。’”

爱因斯坦的答案是：没有绝对空间那样的东西，也没有绝对时间那样的东西。牛顿的物理学基础完全崩溃了。至于以太，那是不存在的。

爱因斯坦抛弃了绝对空间，“在绝对空间中静止”的说法就绝对没有意义了。他声称，没有办法测量地球在绝对空间的运动，这就是为什么迈克尔逊-莫雷实验会出现那样的结果。我们只能测量地球相对于其他自然事物（如太阳、月亮）的速度，正如我们也只能测量火车相对于大地、空气等自然物的速度一样。不论地球、火车还是别的任何事物，都没有绝对运动的依据；运动纯粹是“相对的”。

爱因斯坦抛弃了绝对空间，也就抛弃了这样的观念：不论如何运动，关于某张桌子、某列火车或某个别的什么东西，每个人都能得到一样的长、宽和高。相反，爱因斯坦坚信，长、宽、高都是“相对的”概念，它们依赖于被测物体和测量者的相对运动。

爱因斯坦抛弃了绝对时间，也就抛弃了不论如何运动，每个人都得以相同方式经历时间流的观念。爱因斯坦宣布，时间是相对的。每

一个以自己方式旅行的人，一定会与其他以不同方式旅行的人，经历不同的时间流。

在这些论断面前，我们难免会感到不安。如果它们是正确的，那么它们不但会破坏整个牛顿物理学定律大厦的基础，而且还将剥夺我们的普通感觉，变革我们对空间和时间的日常观念。

但是，爱因斯坦不仅是破坏者，也是创造者。他为我们提出了一个取代旧基础的新基础，这个基础当然是牢固的，而且已经证明，它同宇宙的和谐要完美得多。

爱因斯坦的新基础由两个新的基本原理构成：

·光速绝对性原理：不论空间和时间的本性如何，它们的构成必定使光速在所有方向上都绝对地相同，而且绝对与测量者的运动无关。

这个原理响亮地宣布了，迈克尔逊-莫雷实验是正确的。而且，不论未来的测量装置多么精确，它们一定会得出相同的结果：一个普适的光速。

·相对性原理：不论物理学定律的本质如何，它们都必须在同等的视点上处理所有的运动状态。

这个原理断然抛弃了绝对空间：如果物理学定律不在同等的视点上讨论所有的运动状态（例如，太阳的运动状态和地球的运动状态），那么利用这些物理定律，物理学家就能选出某个“优越的”运动状态（如太阳的）并将它定义为“绝对静止”状态。这样的话，绝对空间又将溜回物理学。在本章后面，我们还会谈这个问题。

根据光速的绝对性，爱因斯坦用后面卡片1.1中所述的精巧的逻辑论证方法证明了，如果你我彼此相对运动，那么，我所谓的空间一定是你的空间和你的时间的混合，而你所谓的空间一定是我的空间和

我的时间的混合。

这儿说的“空间和时间的混合”，类似于地球上的方向。大自然为我们提供了两种确定方向的办法，一种关系着地球的自转，另一种关系着地磁场。在加利福尼亚的帕萨迪纳，地磁北极（罗盘针所指的方向）与真实北极（通过地球自转轴，即通过“北极星”的方向）偏离了大约20度，见图1.2。这意味着，为了在磁北方向上旅行，我们的路线必须部分（约80%）沿真北方向，部分（约20%）沿真东方向。在这个意义上，磁北是真北与真东的混合。同样，真北也是磁北和磁东的混合。

为了理解类似的空间和时间的混合（你的空间是我的空间和时间的混合，我的空间是你的空间和时间的混合），想象你有辆大马力的赛车，你喜欢在深夜以极高的速度在帕萨迪纳的科罗拉多林阴大道上飞驰，而我是警察，那时正在打瞌睡。你在汽车顶上拴了许多鞭炮，引擎盖前一只，车身后一只，中间还有很多，见图1.3（a）。当你通过我的岗亭时，照你的观察，你同时点燃鞭炮。

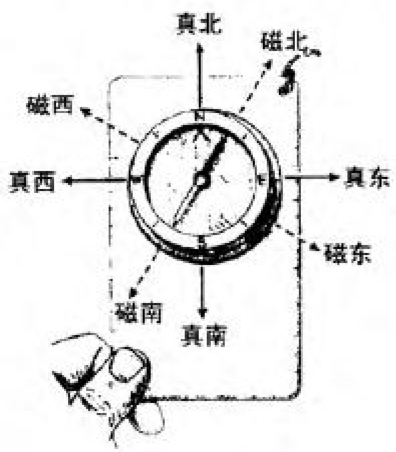



图1.2　磁北是真北和真东的混合，而真北是磁北和磁东的混合

图1.3（b）是照你的视点画的。竖直线是你所测得的时间流（“你的时间”）。水平线是你测量的从车尾到车头的距离（“你的空间”）。因为鞭炮在你的空间里（也就是，照你的观察）是静止的，所以随着你的时间的流逝，它们保持在图中的同一个水平位置上，如虚线所示，每根线代表一只鞭炮。这些线都垂直向上延伸，表明不管时间如何流逝，空间里没有向右或向左的运动——延伸突然终结在鞭炮爆炸的时刻。爆炸的事件在图中以星号表示。

这种图叫做时空图，它以水平方向画空间，以垂直方向画时间；虚线叫世界线，因为它们表示当时间流过时，鞭炮在世界的什么地方运行。以后，我们还会更多地发挥时空图和世界线的作用。

如果谁在图中（图1.3（b））的水平方向上运动，那么他实际上是在你的时间的一个固定时刻通过空间。相应地，我们方便地认为，图中的每一条水平线描述了你在你的时间的某个时刻所看到的空间（“你的空间”）。例如，点画的水平线就是你在鞭炮爆炸时刻的空间。如果谁在图中竖直地向上运动，那么他实际上是在你的空间的一个固定位置上穿过时间。相应地，我们方便地认为，时空图中的每一条竖直线（如每个鞭炮的世界线）描述了你在空间的某个位置上的时间流。

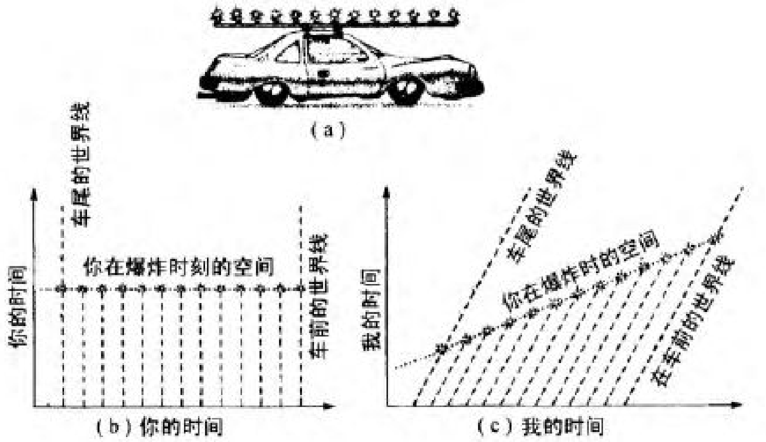



图1.3　（a）你的跑车车顶系着鞭炮在科罗拉多林阴大道上飞驰

（b）根据你的视点（行驶中的汽车）画的鞭炮运动和爆炸的时空图

（c）根据我的视点（静止在岗亭中）画的鞭炮运动和爆炸的时空图

在岗亭里的我，如果没打瞌睡的话，会画一幅很不一样的时空图来描绘你的汽车、你的鞭炮和爆炸（图1.3（c））。我用竖直线画我所测得的时间流，用水平线画沿着科罗拉多林阴大道的距离。随着时间流过，每个鞭炮都跟汽车一起高速地在林阴大道上运动，相应地，鞭炮的世界线在图中向右倾斜：爆炸时，右边的鞭炮比开始时离得更远。

现在，我们来看爱因斯坦逻辑论证（卡片1.1）的惊人结论。光速的绝对性要求，在我看来，鞭炮不会同时爆炸，即使在你看来，它们是同时爆炸的。据我的观点，你车上最边缘的鞭炮最先爆炸，而最前沿的鞭炮最后爆炸。相应的是，我们称为“你在爆炸时刻的空间”的点画线（图1.3（b））在我的时空图中是倾斜的（图1.3（c））。

图1.3（c）清楚地表明，为了在你的爆炸时刻通过你的空间（沿点画的爆炸线），我必须在我的空间和时间中穿过。从这个意义说，你的空间是我的空间和时间的混合，这跟我们说磁北是真北和真东的

混合，是同样的意思（比较图1.3（c）和图1.2）。

你可能忍不住想说，“空间和时间的混合”不过是“同时性依赖于人们的运动状态”的一种复杂和虚张声势的说法而已。是的。不过，在爱因斯坦基础上建设的物理学家们发现，这样的思维方式是很有威力的，它曾帮助他们破译爱因斯坦留下的遗产（他的新物理学定律），在那些遗产中发现了一系列看似古怪的现象：黑洞、虫洞、奇点、时间弯曲和时间机器。

根据相对性原理和光速的绝对性原理，爱因斯坦得到了其他一些空间和时间的显著特征。用上面那个故事的话来说：

·爱因斯坦认为，当你在科罗拉多林阴大道上向东行驶时，我一定会发现你的空间和在其中静止的一切事物（你的车、你的鞭炮和你自己）在东西方向（而不是南北或上下方向）上缩短了，这就是菲兹杰拉德所推测的收缩，不过现在找到了坚实的基础：收缩是由空间和时间的特殊性质引起的，而不是什么作用在运动物体上的自然力的结果。

·同样，爱因斯坦也认为，当你向东运动时，你一定会发现我的空间和在其中静止的一切事物（我的岗亭、我的桌子和我自己）在东西方向（而不是南北方向和上下方向）上收缩了。你看我收缩，我看你收缩，这似乎令人困惑，但实际上不可能再有别的结果：它将你我的运动状态放在了一个平等的基础上，这正符合相对性原理。

·爱因斯坦还认为，在快速驶过时，我发现你的时间流慢了，也就是说，时间膨胀了。你车上仪表板的钟比我岗亭墙上的钟显得要慢些。与我相比，你说话更慢，你的头发长得更慢，你的年岁也过得更慢了。

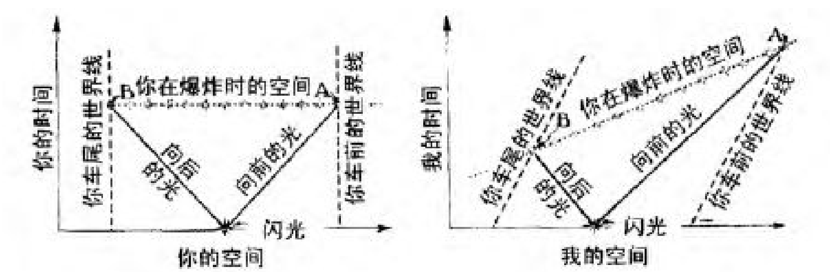
·同样，根据相对性原理，当你从我身边驶过时，你会发现我的时

间流慢了，你看到我岗亭墙上的钟比你仪表板上的钟走得慢。对你来说，我好像也是说话慢了，头发长慢了，我的年岁也过得慢了。

卡片1.1 爱因斯坦对空间和时间混合的证明

爱因斯坦的光速绝对性原理迫使空间和时间相混合，换句话说，它强调同时性是相对的：在你看来同时发生的事件（在你的时间的某一时刻，在你的空间中，如你的赛车在科罗拉多林阴大道上奔驰），在我这个坐在岗亭里的警察看来，并不是同时发生的。我将用与下面所示的时空图相联系的描述性语言来证明这一点，这个证明与爱因斯坦在1905年提出的基本相同。[24]

在你的车中央放一只闪光灯。把灯打开，它向车头发出一道向前的闪光，向车尾发出一道向后的闪光。由于两个闪光是同时发出的，由于根据你在车中的测量，它们经过相同的距离，由于它们以相同的速度传播（光速是绝对的），因此在你看来，它们一定同时到达车头和车尾，看下面左边的图。于是，根据你的观点，两个闪光事件（在车头的为A，车尾的为B）是同时发生的，而且刚好与你看到的图1.3中的鞭炮爆炸相吻合。




接下来让我们看看，从我的视点观察，当你的汽车快速从我眼前驶过时，两个闪光和它们到达的事件A、B是如何的，请看上面右边的图。据我的观点，你的车尾在向前运动，接近向后传来的闪光，所以，我看到它们相遇（事件B）比你看到的更早。同样，你的车头在

向前运动，但远离向前的闪光，所以，我看到它们相遇（事件A）比你看到的更晚（这个结论的关键在于这样的事实：我所看到的两个闪光的速度是相同的；也就是说，结论依赖于光速的绝对性）。因此，我认为事件B发生在事件A的前头；同样，我看到靠近车尾的鞭炮比靠近车头的鞭炮先爆炸。

注意，爆炸在上面的时空图中的位置（在你的时间的某一时刻你所在的空间）与图1.3是一样的。这证明了我们将在下面讨论的空间和时间的混合。

我看你的时间流慢了，而你看我的时间流慢了，这怎么可能呢？这是什么逻辑啊？另外，我怎么能看到你的空间收缩了，而你看我的空间也收缩了？答案都依赖于同时的相对性。关于在我们各自空间中不同位置发生的事件是否同时，你和我没有一致的结论，而这种不协调看来正好协调了我们在时间流和空间收缩上的矛盾，它也靠这个方式保证了一切事物在逻辑上的一致。不过，为说明这种逻辑的一致，需要花太多的篇幅，我不想那么做，请你去看泰勒（Taylor）和惠勒（Wheeler）在1992年的那本书的第3章里的证明。

我们人类在日常生活中从来没有感觉到空间和时间的这类怪异行为，那是怎么回事呢？答案是，我们的运动太慢了。我们相对于彼此的运动速度，总是远远小于光速（每秒299　792公里）。假如你的车在科罗拉多林阴大道上以每小时150公里的速度疾驰，那么我知道你的时间流膨胀和你的空间收缩的量大约是一百万亿分之一（$1\times10^{-14}$），这对我们的感觉来说是太小了。不过，如果你的车以光速的87%的速度冲过我，那么，我（用反应极快的仪器）可以发现，你的时间流比我的慢2倍，而你也看到我的时间流比你的慢2倍；同样，我会看到你车上的所有物体在东西方向上的长度都只有正常情况下的一半；而你也会看到，我岗亭里的所有物体在东西方向上的长度也只有正常情况下的一半。实际上，20世纪后期的大量实验都证实了，空间和时间正是以这种方式发生作用的。[25]

爱因斯坦是如何得到这些空间和时间的基本描述的呢？

他没有检验什么实验结果。在他那个年代，只有低速运动，钟也不够精确，不可能表现出任何时间的膨胀和同时性的不一致，而且量杆也不够精确，表现不出长度的收缩。那时，相关的实验也只有那么几个，如迈克尔逊和莫雷证实地面光速可能在各个方向都相同的实验。这些数据对建立这样一个关于空间和时间的概念基础，是远远不够的！而且，爱因斯坦对这些实验几乎没怎么留意。

实际上，爱因斯坦是靠他天生的直觉来判断哪些事情是应该相信的。经过反复的思考，光速一定是一个独立于方向、独立于运动的普适常数，在他的直觉看来，成了显然的事实。他推论，只有在这种情况下，麦克斯韦的电磁学定律才会始终是简单而优美的（例如，“磁力线永远没有端点”），而且他坚信，在某个深层的意义上，宇宙也愿意拥有简单而优美的定律。就这样，他引进了一个新原理，他的光速绝对性原理，作为一切物理学的基础。

凭这个原理本身，不需要别的东西，就已经确保了建立在爱因斯坦基础上的物理学定律的大厦将完全不同于牛顿的。牛顿物理学家假定空间和时间是绝对的，他必然得到的结论是，光速是相对的——它依赖于事物的运动状态（像本章先前提到的鸟和火车的类比那样）。爱因斯坦假定光速是绝对的，他必然得到的结论是，空间和时间是相对的——它依赖于事物的运动状态。在得到空间和时间是相对的结论后，对简单和优美的追求又将爱因斯坦引向他的相对性原理：没有哪个运动状态会比其他状态更优越，在物理学定律看来，一切运动状态都是平等的。[26]

对爱因斯坦的物理学新基础的构建来说，不仅实验不重要，其他物理学家的思想也不重要。他几乎不关心别人在做什么，他甚至连洛伦兹、庞卡莱、拉莫和其他作者在1896年到1905年间所写的关于空间、时间和以太的那些重要的专业论文也没读过一篇。

洛伦兹、彭加勒和拉莫也在他们的文章里摸索爱因斯坦那样的对我们空间和时间概念的修正，但他们却迷失在牛顿物理学强加给他们的错误概念的迷雾中，而爱因斯坦却能够完全抛弃这些错误概念。他相信宇宙喜欢简单和优美；他情愿追随这个信念，尽管它意味着破坏牛顿物理学的基础。这为他带来了无比清晰的空间和时间的新图景。

在本书后面，相对性原理还会发挥重要作用，所以我再用几页来更深入地解释一下。

我们首先应该有一个参照系的概念。一个参照系就是一个实验室，它有各种测量仪器，你可以在里面做任何你想做的测量。实验室跟它所有仪器一道在宇宙中运动，它们必须经历相同的运动。事实上，参照系的运动才是真正重要的概念，当代物理学家说“不同的参照系”时，他强调的正是两个实验室的不同运动状态，而不是不同的测量仪器。

参照系的实验室和仪器都不必是真的，它们完全可以是一种想象的结构，只存在于物理学家的头脑中，使他们可以问某个问题，例如，“假如我在小行星带中穿行的宇宙飞船里，我想测量某颗小行星的大小，结果会怎样？”这些物理学家实际是想象他们有一个固定在宇宙飞船上的参照系（实验室），想象他们可以用那个参照系的仪器来进行测量。

爱因斯坦在表述他的相对性原理时，没有用任意的参照系，而是用了一类相当特殊的参照系：一类既不自己加速也不受外力推动而只靠自身惯性自由运动的参照系（实验室），因而它总是保持它开始的那种匀速运动的状态。爱因斯坦称这类参照系为惯性系，因为它们的运动完全受惯性的支配。

固定在点火的火箭上的参照系（火箭里的实验室）就不是惯性系，因为它的运动不仅受惯性作用，还受火箭推进的影响。推进的系

统的运动不再是匀速的了。固定在航天器上的参照系，在重新进入地球大气层时，也不是惯性的，因为航天器表面与地球空气分子的摩擦会使它变慢，从而运动不再是匀速的。

最重要的是，在任何大质量物体（如地球）附近，所有参照系都受引力作用，没有什么办法可以让参照系（或其他事物）躲避引力的吸引。所以，在惯性系的局限下，爱因斯坦在1905年无法考虑当引力起重要作用时的物理状态。[1]实际上，他将我们的宇宙理想化为一个完全没有引力的世界。像这类极端的理想化对物理学的进步是很重要的。我们从概念上放弃宇宙的难以理解的方面，而当我们理性地把握了它的其余方面，即相对较容易的方面后，再回到那些困难的方面来。爱因斯坦在1905年理性地把握了没有引力的理想化的宇宙，然后他才转向一个更困难的任务，去认识真实的由引力主宰的宇宙的空间和时间的本性，这个任务最终迫使他得到这样的结论：引力卷曲了空间和时间（第2章）。

理解了惯性参照系的概念，我们现在更深入地来讨论爱因斯坦相对性原理更准确的形式：以在一个惯性系中所进行的测量来建立任何物理学定律，那么，当以在任何其他惯性系中所进行的测量来重建这些定律时，它们必须具有与在原来参照系中完全相同的数学形式和逻辑形式。换句话说，物理学定律不必为我们提供区别一个惯性系（一种匀速运动状态）和任何其他惯性系的方法。

看两个物理定律的例子，会更明白这一点：

·“初始静止在惯性参照系中的任意自由物体（即不受力作用的物体）将总保持静止；原来在惯性参照系中运动的任何自由物体，将永远以不变的速度沿直线向前运动。”如果（确实如此）我们有足够的理由相信，这个牛顿第一运动定律的相对论表述至少在一个惯性系中是正确的，那么，相对性原理认定，它在一切惯性参照系中也一定是正确的，不管这些参照系在宇宙的什么地方，也不管它们运动得多么

快。

·麦克斯韦的电磁学定律必须在所有参照系中具有相同的数学形式。当我们在牛顿基础上建立这些定律时，它们并不如此（磁力线在某些参照系中可能有端点，而在另一些参照系中却没有端点），这一缺陷深深刺激了洛伦兹、彭加勒、拉莫和爱因斯坦。在爱因斯坦看来，这些定律在一个参照系，即以太所在的参照系中简单而优美，但在其他所有相对于以太运动的参照系中却复杂而丑陋，这是不能接受的。通过重建物理学基础，爱因斯坦也使麦克斯韦定律在各惯性参照系中都有了一个简单而优美的形式，而在每个惯性系中的形式都是一样的（例如，“磁力线永远不会有端点”）——这正符合他的相对性原理。

相对性原理实际上是一个形而上的原理（metaprinciple），原因是，它本身并不是一个物理学定律，而是一种模式或规则，（爱因斯坦断言）所有的物理学定律都必须遵从这个原理，不论它是什么样的定律，也不论它是关于电的和磁的，或原子的和分子的，还是蒸汽机的和赛车的。这个形而上原理的力量是惊人的。每个新提出的定律都得经受它的检验。如果新定律通过了它的检验（即如果定律在每个惯性系中是一样的），那么这个定律也就有希望描绘我们的宇宙行为。如果检验失败了，那么爱因斯坦会断言它没有希望了，应该被抛弃。

自1905年以来近百年的所有经验告诉我们，爱因斯坦是对的。所有成功描述了真实宇宙的新定律都已证明是服从爱因斯坦的相对性原理的。这个形而上原理，已经成为物理学定律的定律。

1905年5月，爱因斯坦与贝索的讨论帮他打碎了思想上的拦路石，使他抛弃了绝对的时间和空间，接下来，他只思考和计算了几个星期，就形成了他的物理学新基础，导出了一系列关于空间、时间、电磁和高速运动物体行为本质的结论。其中的两个结论是很辉煌的：质量可以转化为能量（这将成为原子弹的基础，见第6章）；每个物

体的惯性在速度接近光速时必然会快速地增大，以至不论我们费多大力量推动它，都不可能使它达到或超过光速（“没有什么能比光还跑得快”）。[2]

6月底，爱因斯坦将他的思想和结论写成一篇论文，投给《物理学纪事》。他的论文有一个寻常意味的题目：“论运动物体的电动力学”，内容却是异乎寻常的。匆匆读过，我们会看到，爱因斯坦，这位瑞士专利局的“三级技术员”，提出了一个全新的物理学基础，提出了一个未来所有的物理学定律都必须遵从的形而上原理，还极大地修正了我们的空间和时间的观念，导出了辉煌的结论。很快，爱因斯坦的新基础和结论就出名了，叫做狭义相对论（说它“狭义”是因为它只是在引力不重要的特殊情况下正确描绘了宇宙）。

莱比锡《物理学纪事》编辑部在1905年6月30日收到爱因斯坦的论文，经过仔细认真的审读，论文通过了，被接受了，发表了。[27]

在论文发表后的几个星期里，爱因斯坦期待着来自当代大物理学家的反应。他的观点和结论太基本了，几乎没有实验基础，所以他等待着尖锐的批评和争论，然而，他等来的是冷漠的沉寂。许多个星期过去了，最后他收到一封来自柏林的信：马克斯·普朗克（Max Planck）询问文章里的几个技术细节，请他说明。爱因斯坦真是欣喜若狂！普朗克是还健在的最有名的物理学家之一，能受到他的注意，实在是令人满足的。第二年，当普朗克继续以爱因斯坦的相对性原理作为自己研究的核心工具时，爱因斯坦更加振奋了。因为普朗克的赞扬，因为慢慢来自其他杰出物理学家的赞扬，而最重要的，因为他本人极端的自信，当他所预料的争论在接下来的20年里真的纷扰在相对论周围时，爱因斯坦还能够坚定地挺过来。直到1922年，争论仍然很激烈，所以当瑞典科学院的秘书电告爱因斯坦获得诺贝尔奖时，还特别指出，相对论不在评奖所考虑的工作之内。

争论到30年代才最后结束，那时技术已经很先进了，可以为狭义

相对论的预言带来精确的实验证据。到90年代的今天，更没有丝毫可以怀疑的了：在斯坦福大学、康奈尔大学和其他地方的粒子加速器里，每天有$10^{17}$以上的电子被加速到高达0.999 999999 5个光速——而它们在这种超高速下的行为完全与爱因斯坦狭义相对论的物理学定律相吻合。例如，随着速度接近光速，电子的惯性增加了，使它不能达到光速；当电子与靶子碰撞时，它们产生高速的被称为μ子的粒子。以μ子自身的时间来测量，它们只能生存2.22微秒，但是以静止在实验室中的物理学家的时间来测量，由于时间膨胀，它们可以生存100微秒或更长的时间。

[1] 这就是说，我在上面例子中用的高速汽车是不太恰当的，它也受地球的引力作用。不过，因为地球引力是垂直作用在汽车的运动方向上的（也就是一个向下，一个水平），它不会给我们在赛车故事里讨论的观点带来任何影响。

[2] 不过这是有条件的，见第14章。

## 物理学定律的本质

爱因斯坦狭义相对论的成功是否意味着我们必须完全抛弃牛顿的物理学定律呢？显然不是。在日常生活里，在大多数科学领域和大多数技术应用中，牛顿定律仍然被广泛运用着。我们在计划乘飞机旅行时不会关心时间膨胀；工程师在设计飞机时也不会为长度收缩而焦虑。这类膨胀和收缩太小了，用不着关心。

当然，如果愿意，我们可以在日常生活中运用爱因斯坦的定律，而不用牛顿的。两者对一切物理效应都给出几乎完全一样的预言，因为日常生活中达到的相对速度同光速比起来真是太小了。

只有在相对速度接近光速时，爱因斯坦和牛顿的预言才开始出现严重的分歧。这时，也只有在这时，我们才必须抛弃牛顿而严格忠实于爱因斯坦的预言。

这是一个极普遍的模式的一个例子，在未来的章节里我们还会遇到。这种模式在20世纪的物理学历史上曾反复出现：一组定律（在我们这儿，即牛顿定律）起初被广泛接受，因为它与实验吻合得很好。但是，随着实验越来越精确，起初的那组定律只有在一定的极限范围，即在定律的有效范围内（对牛顿定律而言，就是速度远小于光速的范围）才能较好地成立。然后，物理学家努力从实验和理论去认识在那个有效范围的边界上发生的事情，最后，他们建立一组在边界内、边界附近和边界以外都高度成功的新定律（在牛顿的情形，爱因斯坦的狭义相对论不仅对低速有效，在近光速时也有效）。物理学定律的历史重复着这个过程，在以后的章节，我们还会遇到这样的重复：当引力变得重要时，狭义相对论将失败，取而代之的是一组叫广义相对论的新定律（第2章）；在黑洞内部奇点的邻近，广义相对论将失败，取而代之的是一组叫量子引力的新定律（第13章）。

从旧定律到新定律的每一次转变，都有一个令人惊讶的特征：在每种情形下，物理学家（如果他们足够聪明）都不需要靠什么实验指引来告诉他们，旧定律会从哪儿开始崩溃，也就是，有效性的边界在哪里。对牛顿物理学来说，我们已经看到了：麦克斯韦的电动力学定律没有很好地与牛顿物理学的绝对空间相吻合。在绝对空间中（即在以太的参照系中）静止时，麦克斯韦的定律简单而优美——例如，磁力线没有端点。在运动参照系中，它们变得复杂而丑陋——磁力线有时有端点。不过，当参照系以远小于光速的速度在绝对空间中运动时，这种复杂对实验结果的影响是可以忽略的，也就是说，几乎所有的力线都没有端点。只有在速度接近光速时，丑陋的复杂性才会带来容易测量的大影响：会出现许多端点。因此，即使没有迈克尔逊-莫雷实验，也有理由相信，牛顿物理学的有效范围是速度远小于光速，而牛顿定律可能会在速度接近光速时崩溃。

类似地，在第2章我们将看到狭义相对论如何预言自己会在引力出现时失败；而在第13章我们将看到广义相对论如何预言自己会在奇点的邻近失败。

在考虑上面那一系列定律（牛顿物理学、狭义相对论、广义相对论、量子引力）以及类似的一系列主宰物质结构和基本粒子的定律时，大多数物理学家都冲动地相信，这些系列的定律将汇聚成一组终极定律，它才真正主宰宇宙，它迫使宇宙照实际的方式运行，迫使雨水在窗户上凝结，迫使太阳燃烧核子，迫使黑洞在碰撞时产生引力波，等等。

可能会有人反驳说，在那个序列中，每一组定律“看起来”都与它前头的那些定律大不相同（例如，牛顿物理学的绝对时间看来就大不同于狭义相对论中的许多时间流）。“看起来”，这些定律没有任何汇聚的征兆。那么，我们为什么还期待着它们的汇聚呢？答案是，我们必须明确地区分一组定律的预言和这些定律所传达的理性图像（定律“像”什么）。我希望的汇聚只是就预言说的，但那也就是最终有意

义的一切。理性的图像（牛顿物理学中的绝对时间，相对论物理学中的许多时间流）对最终的实在的本质来说是不重要的。事实上，我们有可能完全改变一组定律“像”什么，而一点儿也不改变它的预言。在第11章里，我将讨论这个值得注意的事情，会举一些例子，还要解释它对实在的本质有什么意义。

我为什么希望预言意义上的汇聚呢？因为我们所有的证据都指明了这一点。每组定律都比它前头的定律有更大的有效范围：

牛顿定律在日常生活的一切范围内都是成功的，但它不适用于物理学家的粒子加速器，不适用于遥远宇宙的奇异现象，如脉冲星、类星体和黑洞；爱因斯坦的广义相对论定律在我们实验室的各个地方，在遥远宇宙的每一个角落都是成功的，但它在黑洞的深处，在宇宙大爆炸诞生的地方却失败了；量子引力的定律（我们现在还远没有很好地认识）也许会绝对地在任何地方都成功。

在这本书里，我将不加辩解地采纳这个观点：确实存在着一组终极的物理学定律（我们现在还不知道，但也许就是量子引力），它们真正地统治着我们周围的宇宙的各个角落。它们迫使宇宙按它实际的方式运行。如果要说得更准确些，我应该说，我们现在用的定律（如广义相对论）“近似于”真实定律，或者说，它是真实定律的“一种近似描述”。然而，我一般都不提这个限制，也不区分真实定律和我们的近似。在这些情形，我会断言，例如，“广义相对论定律[而不说真实定律]迫使黑洞将光牢牢地抓住，使它不能从黑洞的视界逃脱。”在认识宇宙的奋斗中，我的物理学同行们和我就是这么思考的。这是一种卓有成效的思想方法，为我们带来了关于坍缩的恒星、黑洞、引力波和其他现象的崭新而深刻的认识。

与这种观点对立的是，人们普遍认为，物理学家在同一些理论打交道，这些理论试图描述宇宙，但它们不过是人类的发明，不会对宇宙产生真正的威力。实际上，理论一词包含了太多的试探性和人为的

诡辩意味，我将尽可能回避它。需要的时候，我将在真正主宰宇宙、迫使宇宙以实际方式运行的严格意义上，用物理学定律这个词组来代替它。

# 第2章　空间和时间的卷曲

赫尔曼·闵可夫斯基

统一了空间和时间，

而爱因斯坦令它们发生卷曲

## 闵可夫斯基的绝对时空

我要摆在你们面前的空间和时间的观点，已经从实验物理学的土壤中萌芽了，那里积蓄着它们的力量。它们是基本的。从今往后，空间和时间本身都将注定在黑暗中消失，只有二者的一种结合能保持为一个独立的实体。[1]

1908年9月，赫尔曼·闵可夫斯基用这样的话向世界宣布了关于空间和时间本性的新发现。

爱因斯坦已经证明，空间和时间是“相对的”。物体的长和时间的流从不同参照系看来是不同的。如果我相对于你运动，那么我的时间就不同于你的，我的空间也不同于你的。我的时间是你的时间和空间的混合，我的空间是你的空间和时间的混合。

现在，闵可夫斯基在爱因斯坦工作的基础上发现，宇宙是由一种绝对的而不是相对的四维“时空”结构构成的，这种四维结构在所有参照系看来（当然，我们得学会怎么去“看”）都是一样的，它的存在独立于参照系。

下面的故事（根据泰勒和惠勒1992年的书改编）说明了闵可夫斯基发现的基本思想。

从前，在遥远的东方的大海上有个名叫蒙里迪那的岛，岛上居民有着奇特的风俗和禁忌。每年6月，在一年中最长的那个白天，所有蒙里迪那岛的男人都要乘着一艘大帆船，到遥远的一个叫塞罗那的圣岛去朝觐一只巨大的蟾蜍，蟾蜍将整夜地用恒星和星河、脉冲星和类星体的离奇故事来蛊惑他们。第二天，这些男人会带着神的启示回到蒙里迪那，在未来的一年里，这启示将一直伴随着他们。

每年12月，在一年中最长的那个夜晚，蒙里迪那的女人向塞罗那远航。第二天，她们白天朝觐那只大蟾蜍，夜里回去，满怀着恒星和星河、类星体和脉冲星的幻境。

不过，蒙里迪那的女人绝对不能向岛上的任何一个男人讲她们到塞罗那圣岛的经历，也不能讲蟾蜍告诉她们的任何故事。蒙里迪那的男人也得遵守这个禁令，从不向女人透露他们每年一度的航行。

1905年夏天，蒙里迪那岛一个名叫阿尔伯特的激进青年，他才不管什么文明的禁忌。他发现了两张神圣的地图，并将图泄露给岛上所有的男人和女人。有一张地图是蒙里迪那的女祭司在女人的冬夜远航时用来指引帆船的，另一张是祭司在男人夏日航行时用的。圣图暴露了，岛上的男人是多么羞愧！女人也多么羞愧！但地图摆在那儿，每个人都看到了——太令人吃惊了，塞罗那的位置在两张图上不一样！女人是先向东航行210浪〔1浪＝201.167米〕，然后向北100浪；而男人是先向东航行164.5浪，再向北164.5浪。我们知道，宗教习俗是严厉的，女人和男人都必须在同一个塞罗那圣岛向同一只神圣的蟾蜍乞求每年的灵光。但事情怎么会这样呢？

大多数蒙里迪那人为了遮羞，说暴露的地图是假的。但有一位名叫赫尔曼的聪明老人相信图是真的。他为弄清地图差错的秘密奋斗了

3年。最后，在1908年的一个秋日，真相大白了：原来，蒙里迪那男人的航行用的是磁性罗盘，而女人靠的是恒星（图2.1）。男人通过磁性确定北方和东方，女人则依靠由于地球自转而在头顶旋转的恒星来确定这些方向，两种定向方法偏离20°。当男人向他们确定的北方航行时，在女人看来，他们实际航行在“北偏东20°”的方向上，即约80%的北和20%的东。在这个意义上，男人的北方是女人的北方和东方的混合；同样，女人的北方也是男人的北方和东方的混合。

引导赫尔曼发现这一点的关键是毕达哥拉斯（Pythagoras）公式：取直角三角形的两个腰，将一个腰的平方与另一个腰的平方加起来，取平方根，结果就是三角形斜边的长。

斜边就是从蒙里迪那到塞罗那的直线路径。在女人的地图上，两个腰沿真东和真北方向，照此，沿着这条直线路径的绝对距离是$\sqrt{210^2+100^2}=232.6$（浪）。根据男人的地图，腰在磁东和磁北方向，绝对距离为$\sqrt{164.5^2+164.5^2}=232.6$（浪）。向东和向北的距离是“相对”的，它依赖于地图的参照系是磁方向的还是真方向的。但是，不论根据哪一组相对距离，我们都能计算出同一个绝对的直线距离。

蒙里迪那居民和他们的禁忌文化对这个绝妙的发现有什么反应呢？历史没有记录。

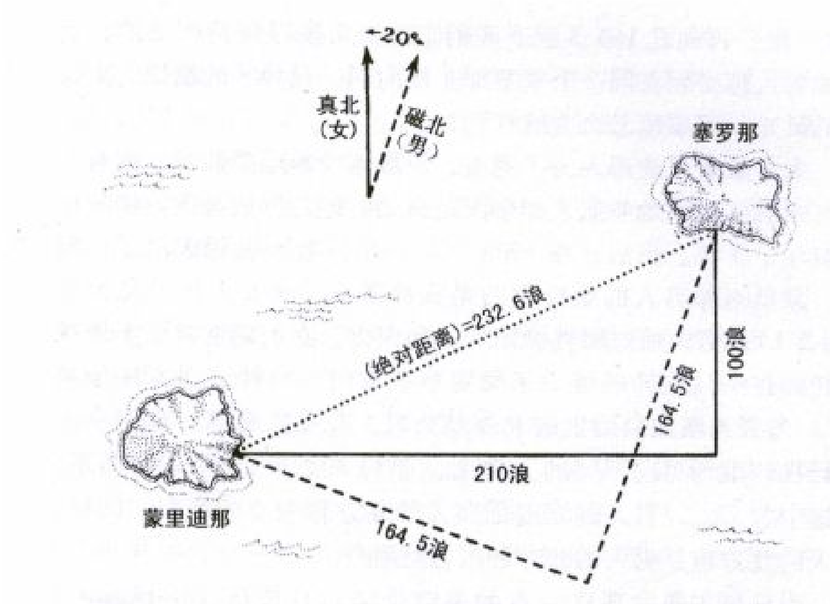



图2.1　两张重叠在一起的从蒙里迪那到塞罗那的路线图，图上有赫尔曼做的磁北、真北和绝对距离的记号

赫尔曼·闵可夫斯基的发现，类似于蒙里迪那的那位赫尔曼老人的发现：假设你相对于我运动（比如，在你超高速的赛车里），那么：

·像磁北是真北和真东的混合一样，我的时间也是你的时间和你的空间的混合。

·像磁东是真东和真北的混合一样，我的空间也是你的空间和你的时间的混合。

·正如磁北和磁东、真北和真东不过是为了在一个先存在的二维曲面，即地球表面上进行测量的不同方式，我的空间和时间，以及你的空间和时间，也不过是为了在一个先存在的被闵可夫斯基称为时空的四维“曲面”或“结构”上进行测量的不同方式。

·正如在地球表面存在一个从蒙里迪那到塞罗那的绝对直线距离

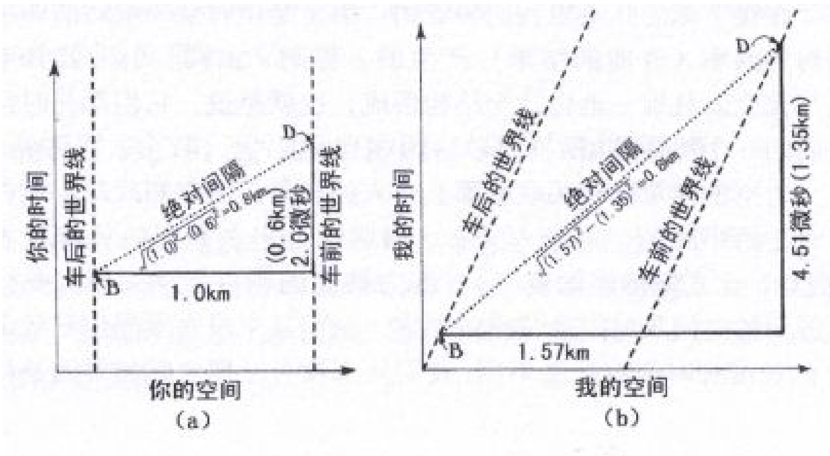
——它可以根据毕达哥拉斯公式，用磁北和磁东方向的距离或用真北和真东方向的距离计算出来——在时空的任意两个事件之间，也存在着一个绝对的直线间隔，它可以根据一个与毕达哥拉斯相类似的公式，用我的或你的参照系中测量的长度和时间计算出来。

闵可夫斯基正是通过与毕达哥拉斯公式的类比（我称它为闵可夫斯基公式），发现了他的绝对时空。

闵可夫斯基公式的细节对本书其余部分是不重要的，我们没有必要掌握它（不过，我还是为好奇的读者在卡片2.1中将它们写出来了）。惟一重要的是，时空的事件类似于空间的点，而且时空中任意两个事件之间存在着一个绝对的间隔，完全类似于一张纸上任意两点间的直线距离。间隔的绝对性（不论用谁的参照系来计算，它的值都是一样的）说明，时空有绝对的实在性，它是一个具有若干与运动无关的性质的四维结构。

卡片2.1　闵可夫斯基公式

你驾着1米长的大马力赛车，以每秒162 000公里的速度（光速的54%）呼啸着从我身边飞过，回想一下图1.3的情形。下面的时空图画出了你的车的运动。图（a）是以你的视点画的，图（b）以我的视点。当你经过我时，汽车回火，从尾气管排出一阵烟，这个回火事件在图中记为B。2微秒（百万分之二秒）后，你看到前面防撞器上的鞭炮爆了，爆炸事件记为D。




因为空间和时间是相对的（你的空间是我的空间和时间的混合），所以，关于回火事件B和爆炸事件D之间的时间间隔，你和我有不同的意见。照你的时间，它们间隔2.0微秒，而在我看来，是4.51微秒。同样，关于事件的空间间隔我们的意见也不同，在你的空间中，是1.0公里，而在我的空间中是1.57公里。尽管有时间和空间上的分歧，我们都同意，两个分离的事件在四维时空里由一条直线联系着，而且我们一致认为，沿这条直线的“绝对间隔”（线的时空长度）是0.8公里。（这类似于蒙里迪那岛的男人和女人们在蒙里迪那和塞罗那之间的直线距离上达成一致。）

我们可以用闵可夫斯基的公式来计算绝对间隔：将事件的时间间隔乘上光速（每秒299792公里），得到图中所示的四舍五入的数（你的为0.60公里，我的为1.35公里）。然后，将事件的时间间隔和空间间隔平方，从平方的空间间隔中减去平方的时间间隔，再取平方根。（这类似于蒙里迪那人东方和北方的距离平方，加起来，然后取平方根。）从图中可以看到，尽管你的时间和空间间隔不同于我的，关于绝对间隔，我们还是得到了相同的答案：0.8公里。

你和我遵从的闵可夫斯基公式与蒙里迪那人遵从的毕达哥拉斯公式之间，只有一点重要的差别：我们的平方间隔是相减而不是相加。

这里的减法是同你正在探索的时空与蒙里迪那人所经历的地球表面的物理学差异密切联系着的——不过，不怕你生气，我不想解释这种联系了，你可以去看泰勒和惠勒（1992）的讨论。

在接下来的几页里我们将看到，引力是由时空的绝对的四维结构的曲率（卷曲的结果）产生的，黑洞、虫洞、引力波和奇点都完全而且惟一地由这个结构形成，也就是说，它们都是时空卷曲的一个特殊类型。

时空的绝对结构关联着那么迷人的现象，而你和我却不能在日常生活中经历，真令人灰心。问题还是出在我们的低速技术（例如，比光慢得多的赛车）。因为彼此的相对运动太慢，我们所经历的空间和时间是分离的两家，我们从来没有发现你和我测量的长度和时间有什么不同（我们从来没有发现空间和时间是相对的），也从来没有发现我们相对的空间和时间统一形成了一个绝对的四维时空结构。

你可能记得，闵可夫斯基就是在爱因斯坦读书时叫他懒狗的那位数学教授。1902年，俄国出身的闵可夫斯基离开了苏黎世ETH，到德国哥廷根（它那时跟现在一样有国际声誉）去担任更有吸引力的教授。在哥廷根，闵可夫斯基研究了爱因斯坦关于狭义相对论的论文，印象很深，这引导他发现了四维时空的绝对性质。

爱因斯坦听说闵可夫斯基的发现时，并不在意。闵可夫斯基只是用一种新的更数学化的语言重写了狭义相对论的定律，而对爱因斯坦来说，数学掩盖了定律背后的物理意义。因为闵可夫斯基不断宣扬他的时空观如何美妙，爱因斯坦开始笑话哥廷根的数学家：他们用那么复杂的语言来描述相对论，物理学家简直弄不懂了。

事实上，笑话落到了爱因斯坦自己身上。在4年后的1912年，他将认识到，为了在狭义相对论中纳入引力，闵可夫斯基的绝对时空是根本性的基础。遗憾的是，闵可夫斯基没能活着看到这一点。1909

年，他死于阑尾炎，那年他45岁。

在本章后面，我还会回来谈闵可夫斯基的绝对时空。不过现在，我得先引出我的故事的另一条线索：牛顿的引力定律和爱因斯坦为了协调它与狭义相对论而迈出的第一步，这是走在他借鉴闵可夫斯基成果之前的一步。

## 牛顿的引力定律，爱因斯坦协调它与相对论的第一步

牛顿将引力想象为一种作用在宇宙中每一对物体间的力，一种将物体相互拉近的力。物体的质量越大、距离越近，这个力就越强。更精确地说，这个力正比于物体质量的乘积，反比于它们之间的距离的平方。

这个引力定律是理性的巨大胜利。它与牛顿的运动定律结合，解释了行星绕太阳的轨道，卫星绕行星的轨道，海洋潮汐的涨落和岩石的崩落；让牛顿和他17世纪的同胞们学会了如何去称量太阳和地球。[1]

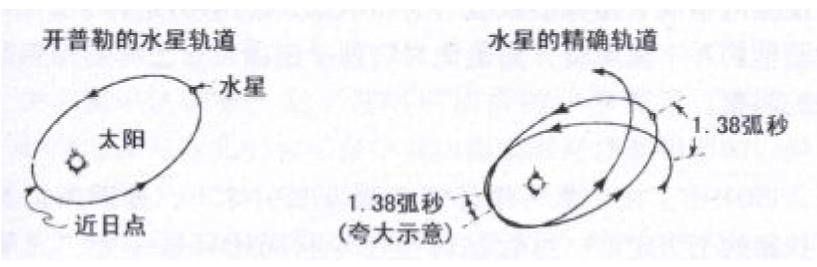
在从牛顿到爱因斯坦之间的两个世纪中，天文学家对天体轨道的测量有了多方面的进步，牛顿的引力定律经受了越来越严格的检验。偶尔会出现一些新的天文测量不符合牛顿定律，但最终也发现这些观测或对它们的解释是错误的。牛顿定律一次又一次地战胜了实验或理性的错误。例如，当天王星（1781年发现）的运动似乎违背了牛顿引力定律的预言时，人们猜想，很可能是因为一颗尚未发现的行星的引力作用在天王星上，干扰了它的轨道。完全依据牛顿的引力和运动定律以及对天王星的观测所进行的计算，预言了新行星应该在天空的某个地方。1846年，当勒维耶（U.J.J.Leverrier）将他的望远镜瞄准那个位置时，预言的行星果然在那儿出现了，尽管对肉眼而言太模糊，用望远镜看却很光亮。这颗捍卫牛顿定律的行星被命名为“海王星”。

20世纪初，牛顿的引力定律还有两个小小的却令人困惑的矛盾。一个是水星轨道的古怪行为，这最终预示了牛顿定律的失败；另一个是月球轨道的异常，后来发现这是天文学家对测量的解释错了。[2]跟精确测量的通常情形一样，很难在这两个矛盾中判别应该忧虑哪一个。

爱因斯坦正确地猜想，水星的古怪行为（它的近日点的反常移动，见卡片2.2）是真的，而月亮的异常不是真的。水星的古怪“闻起来”是真的，而月亮不是。然而，对爱因斯坦来说，实验与引力定律的这个可疑的矛盾并没有多大意思，也不太重要；他相信，更重要也更有意思的是，牛顿定律将违反他新建立的相对性原理（即那个要求一切物理学定律在每个惯性参照系中必须相同的“形而上原理”）。由于爱因斯坦坚信他的相对性原理，所以牛顿定律如果违反了它，就意味着有问题。[2]

卡片2.2　水星近日点的移动

开普勒（Kepler）曾将水星轨道描绘成以太阳为一个焦点的椭圆（左图，轨道椭圆被拉长了）。然而，19世纪的天文学家根据观测发现，水星轨道并不完全是椭圆。水星每沿轨道绕一圈，都不能回到同一个出发点，而是有一点小小的偏离，可以描述为一种移动，即每个轨道在水星离太阳最近的位置发生了移动（轨道的近日点的移动）。天文学家观测到每个轨道的近日点一次移动1.38弧秒（右图，移动被夸大了）。




牛顿的引力定律可以解释这1.38弧秒中的1.28弧秒；那是木星和其他行星对水星的引力作用产生的结果。但是，还剩下0.10弧秒的偏差：水星近日点在每个轨道周期中的0.10弧秒的异常移动。天文学家称，他们的测量误差和不确定程度只有0.01弧秒的大小，但考虑到所测角度太小（0.01弧秒相当于人的一根头发的直径在10公里距离处所

张的角），我们一点儿也不会奇怪，19世纪末和20世纪初的许多物理学家还会对此表示怀疑，并且期待着牛顿定律的最后胜利。

爱因斯坦的理由很简单：照牛顿的观点，引力依赖于两个吸引物体（如太阳和水星）之间的距离，但根据相对论，这个距离在不同参照系中是不同的。例如，爱因斯坦的相对论定律预言，太阳与水星间的距离依赖于我们是在水星表面测量还是在太阳表面测量，两者会产生大约十亿分之一的差别。如果水星和太阳的这两个参照系在物理学定律看来都一样好，那么应该用哪个参照系来测量出现在牛顿引力定律中的距离呢？不论选择水星的还是太阳的参照系，都会违反相对性原理。这种进退两难的境地，使爱因斯坦确信，牛顿的引力定律一定有问题。

爱因斯坦的胆识令人惊讶。他已经在几乎没有实验证据的情况下抛弃了牛顿的绝对空间和绝对时间，现在他又要在更缺少实验证据的情况下抛弃牛顿获得过巨大成功的引力定律了。不过，激励他的并不是实验，而是他对物理学定律应该怎样的深刻的直觉的洞察。

1907年，在一个写作计划的激发和引导下，爱因斯坦开始寻找新的引力定律。尽管这时他在专利局还只是一个“二级技术员”（刚从三级提升的），但全世界的大物理学家都很尊重他，所以有人请他为年刊《放射学与电子学年鉴》写一篇关于他的相对论物理学定律及其结果的综述。[3]爱因斯坦在写作时发现了一条对科学研究很有价值的思路：当我们要把一个主题以一种自洽的、一致的、适于教学的方式向公众展开时，我们被迫以新的方式来思考这个题目，被迫去考察它的所有缺陷和问题，并找寻弥补的办法。

在他的主题中，引力是最大的缺陷。狭义相对论和它不受引力作用的惯性系完全忽略了引力的作用。所以，爱因斯坦在写作中，一直在寻找将引力纳人他的相对论定律的途径。像大多数被问题困惑的人一样，即使在没有直接考虑这个问题时，他的内心也还在想着它。于

是，在1907年11月的某一天，用爱因斯坦自己的话说，“我正坐在伯尔尼专利局的桌旁时，突然出现一个想法：‘如果一个人自由下落，他将感觉不到自己的重量。’”

你我今天也能有这种想法，但引不出什么结果。爱因斯坦却不同，他会追到思想的尽头，向它们索求每一点灵感。落体的想法是关键的，它指向了引力的革命性的新观点。他后来说它是“我一生中最快乐的思想”。

这个思想的结论滚滚而来，成为爱因斯坦那篇综述中的不朽篇章。假如你自由落下（如从悬崖上跳下），你不仅感觉不到自己的重量，而且还会在所有方面都感到，似乎引力完全从你的邻近消失了。例如，你在下落时从手上放落一些石块，你和石块将肩并肩地下落。如果你看着石块而忽略周围的其他事物，你不能判断自己和石块是在向着地面落下，还是远离引力物而在空中自由漂浮。事实上，在你的邻近，引力是没有作用的，不可能观测到。爱因斯坦认识到，在下落时所携带的小参照系（实验室）里，物理定律与在无引力宇宙中自由运动时必须是相同的。换句话说，你自由下落的小参照系“等效于”无引力宇宙中的惯性参照系，你所经历的物理学定律与在无引力惯性系中的是一样的，它们也就是狭义相对论的定律。（以后我们将知道，为什么参照系必须是小的，“小”的意思是，与地球的大小相比，它很小——或者，更一般地说，与引力在强度和方向上发生改变的范围相比，它很小。）

我们来看一个无引力惯性系与自由下落的小参照系等效的例子，考虑在无引力宇宙中自由运动物体（假定它是一颗炮弹）行为的狭义相对论定律。从那个理想化宇宙中的任何惯性系看，炮弹一定沿直线以均匀速度运动。现在将它与在我们真实的引力宇宙中的运动进行比较：如果炮弹从地球的草地上的大炮中发射出来，从坐在草地上的一只狗来看，它将沿弧线向上，飞到空中，然后落回地球（图2.2）。在狗的参照系中，它沿一条抛物线（黑实线）运动。爱因斯坦请你在

一个自由下落的小参照系中观察同一颗炮弹，如果草地有一个悬崖的边缘，这是很容易做到的。你可以在大炮发射时从悬崖跳下去，一边下落一边观察。

为了帮你描绘你下落时所看到的景象，想象你在面前举着一扇有12格玻璃的窗户，你透过玻璃观察炮弹（图2.2中间）。在下落中你会看到像图2.2画的顺时针图像序列。在看这个序列时，要忽略狗、大炮、树木和悬崖，只注意你的窗户格子和炮弹。在你看来，炮弹相对于你的窗户格子以不变的速度沿点画的直线运动。

这样，在狗的参照系里，炮弹服从牛顿定律，沿抛物线运动。在你自由下落的小参照系里，炮弹服从无引力的狭义相对论定律，沿直线匀速运动。而在这个例子中真实的事情在一般情况下也应该是真实的，从这个思想迈出一大步，爱因斯坦认识到：在我们真实的引力宇宙的任何地方的任何自由下落的小参照系中，物理学定律必须与它们在理想化的无引力宇宙的惯性参照系中相同。爱因斯坦称它为等效原理，因为它断言，在引力存在时自由下落的小参照系与没有引力的惯性系是等效的。

爱因斯坦发现，这一断言有一个极其重要的结论：它意味着，只要我们把真实的引力宇宙中的每一个自由下落的小参照系（例如，你从悬崖上落下时带着的小实验室）都称做“惯性参照系”，那么，狭义相对论在理想的无引力宇宙中的惯性系的一切结果，在真实的宇宙中自然也将是正确的。最重要的是，相对性原理必须正确：我们真实的引力宇宙中的惯性的（自由下落的）小参照系必须“构造成为等效的”，在物理学定律看来，没有哪个参照系会比其他任何一个更优越，或者，我们可以更准确地说（见第1章）：

以在一个惯性的（自由下落的）小参照系中所进行的测量来建立任何物理学定律，那么当以任何其他惯性的（自由下落的）小参照系中所进行的测量来重建这些定律时，它们必须具有与在原来的参照系

中完全相同的数学形式和逻辑形式。而且，不论（自由下落的）惯性系是在无引力的星际空间，或者是从地球的悬崖上落下，或者处在我们的银河中心，或者落下来穿过黑洞的视界，它都是正确的。

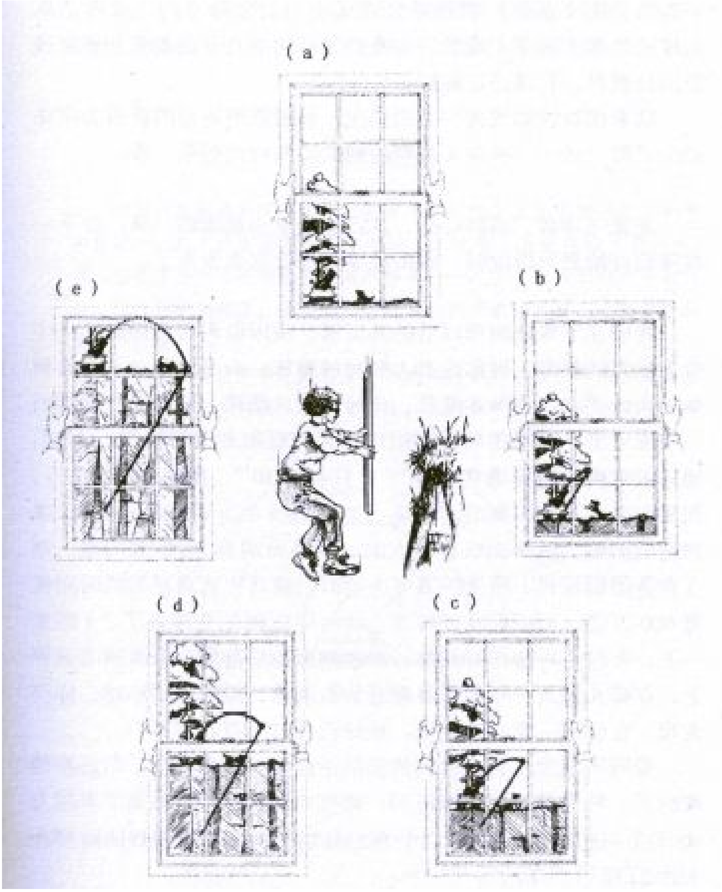



图2.2　中心：你面前举着带十二个格子的窗户从悬崖上跳下。其余的图，从顶上一幅起，依顺时针方向，是大炮发射时你透过窗户看到的情景。相对于下落的窗户参照系，炮弹的轨迹是点画的直线；相对于狗和地球表面，轨迹是实抛物线。

随着相对性原理向引力的扩张，爱因斯坦向他的新引力定律迈出了第一步——从狭义相对论到广义相对论的第一步。

亲爱的读者，请耐心些，这可能是全书最难的一章。在下一章我

们开始黑洞历险时，我的故事就不会这么专业了。

在建立了等效原理以后的几天里，爱因斯坦用它得到了一个令人惊愕的预言，被称为引力的时间膨胀：如果谁相对于引力物体静止，那么，离物体越近，他的时间流越慢。例如，在地球的一间屋子里，时间在地板附近比在天花板附近流得更慢。不过，地球上的快慢差异确实太小了（只有$3/10^{16}$，即亿亿分之三），探测起来是极端困难的。相反（如我们将在下一章看到的），黑洞附近的引力时间膨胀是巨大的；如果黑洞有10个太阳重，那么在离黑洞视界1厘米的高度上的时间流将比远离视界的时间流慢600万倍，而刚好在视界面上的时间流则完全停止了。（想象一下，有没有可能作时间旅行：假如你正好落到一个黑洞的视界上，在那儿经历一年的视界附近的时间流，然后返回地球，你将发现，在你那一年的时间里，地球已经过千百万年了！）

爱因斯坦发现引力时间膨胀的论证多少有些复杂，但后来他找到了一种简单而优美的证明，漂亮地体现了他的物理学思想方法。这一证明在卡片2.4，[4]它所依赖的光的多普勒频移的解释在卡片2.3。

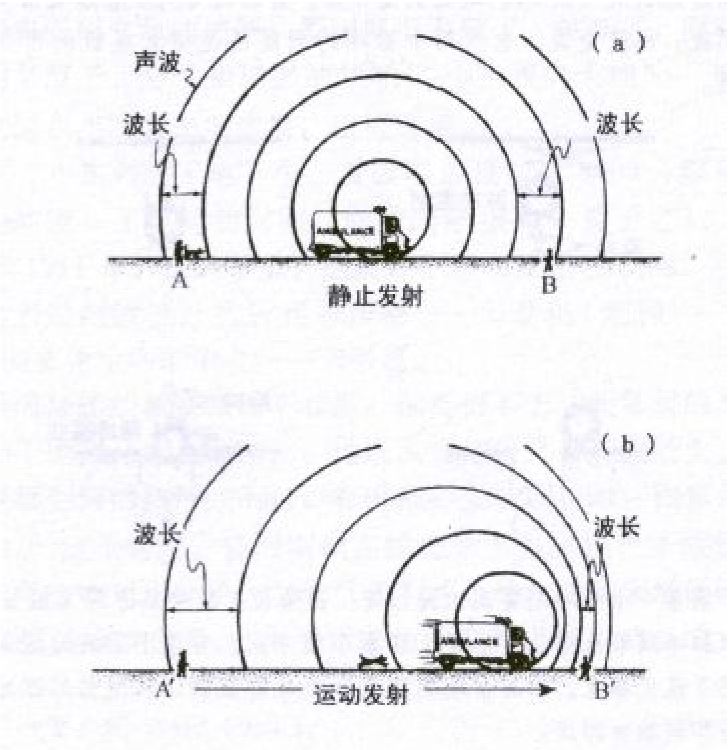
开始写1907年的综述时，爱因斯坦希望它描述无引力宇宙的相对论，但在写作过程中，他发现了三条线索，可能会使引力与他的相对论相吻合——等效原理、引力时间膨胀和他的相对性原理向引力的扩张——所以，他把这些线索也写进去了。大概在12月初，他把文章寄给了《放射学与电子学年鉴》编辑，然后，全身心地去迎接为引力找一个完全的相对论描述的挑战。[5]

卡片2.3　多普勒频移

当波的发射者和接收者相互靠近时，接收者会发现波向更高的频率移动——即更短的周期和更短的波长。如果发射者和接收者分离，那么接收者会发现波向更低的频率移动——即更长的周期和更长的波

长。这叫多普勒频移，是一切类型的波，如声波、水波、电磁波等都具有的性质。

声波的多普勒频移是我们日常熟悉的现象。当救护车尖啸着高速驶过或即将着陆的飞机从头顶飞过时，我们会听到声音突然降低（图b）。想想下面的图，你可能就会理解多普勒频移。



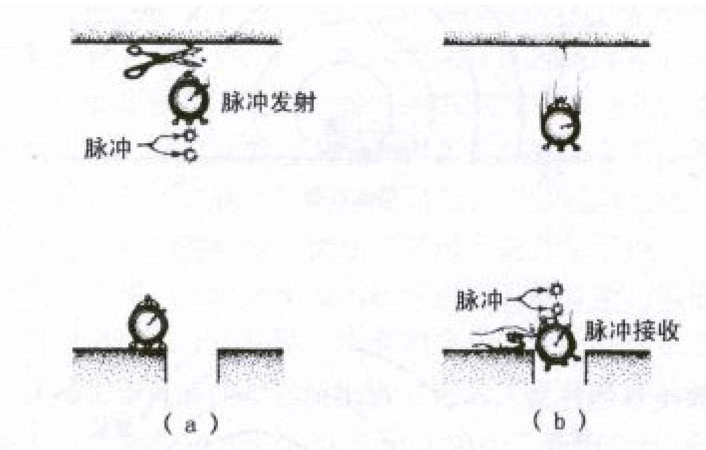

波经历的事情，对脉冲也是正确的。如果发射者发出规则间隔的光（或其他）脉冲，那么，当发射者靠近时，接收者会遇到比在发射时具有更高频率的脉冲（两次脉冲间的时间更短）。

卡片2.4　引力时间膨胀

拿两个相同的钟，一个放在地板上一个洞的旁边（以后钟将落进

这个洞），另一个用绳子吊在天花板上。地板钟的嘀嗒由地板附近的时间流决定，而天花板钟的嘀嗒由天花板附近的时间流决定。

每嘀嗒一声，天花板的钟就发出一个极短的光脉冲，指向下面地板上的钟。在天花板的钟刚要发射第一个脉冲前，将吊它的绳子剪断，让它自由下落。假如嘀嗒声的间隔极短，那么在下一声嘀嗒响起并发射第二个脉冲的时刻，还觉察不到钟的下落，相对于天花板它几乎还处在静止状态（图a）。这必然意味着，钟仍然与天花板感受着相同的时间流，也就是说，它的两个脉冲的间隔还是由天花板的时间流决定的。




当第一个脉冲刚要到达地板时，让地板上的钟落进洞（图b）。第二个脉冲跟着也很快到达了，在两个脉冲间，自由下落的地板钟的运动还不能被察觉，相对于地板它几乎还是静止的，从而仍然跟地板感受着相同的时间流。

通过这样的方式，爱因斯坦将比较天花板和地板所感觉的时间流的问题，转化成为比较两个自由下落的钟（感觉天花板时间的下落的天花板钟与感觉地板时间的下落的地板钟）的嘀嗒速度的问题。然后等效原理又让他可以借助狭义相对论定律来比较两个自由下落的钟。

天花板上的钟因为比地板上的钟先落，它向下的速度总比地板钟的大（图b），也就是说，它在向地板钟靠近。这意味着，地板钟将看到天花板钟发出的经历了多普勒频移（卡片2.3）的光脉冲，即它看到的脉冲到达的时间间隔比它自己嘀嗒的时间间隔更短。由于脉冲间的时间是天花板的时间流决定的，而地板钟的嘀嗒是地板的时间流决定的，这就意味着，地板附近的时间流一定比天花板附近的时间流更慢，换句话说，引力必然使时间流发生膨胀。

12月24日，他给朋友写信说，“我现在正忙着考虑与引力定律相联系的相对论……我希望弄清至今还没能解释的水星近日点移动的长期变化……但似乎还没得到什么结果。”到1908年初，还是没有任何实际的进展，爱因斯坦失望了，放弃了，而将注意力转到了原子、分子和辐射的领域（“小东西的天地”），那里的未解之谜在当时看来更容易，也更有趣。[3]

在“小东西的天地”里，爱因斯坦度过了1908（那年，闵可夫斯基统一了空间和时间，而爱因斯坦却一笑了之），1909，1910和1911年。这些年里，他离开了伯尔尼的专利局，在苏黎世大学当过副教授，然后到布拉格——那是奥（地利）-匈（牙利）帝国文化生活的中心——当教授。

爱因斯坦的教授做得不容易。他不得不上一些常规的与他研究不相干的课，这令他恼火。他既没有把讲义备好的劲头，也没有让课程生辉的热情。不过，在讲他心爱的题目时，他却是精彩绝伦的。[6]这个时候，爱因斯坦在欧洲学术界已经完全成熟起来了，但他也在付出代价。尽管代价不小，他在微观领域的研究却在令人瞩目地推进着，产生了后来为他赢得诺贝尔奖金的大发现（见卡片4.1）。

后来，1911年中期，爱因斯坦对微观的兴趣消退了，他又将精力转向引力的战场，几乎把全部时间都用上了，到1915年11月，他终于成功地建立了广义相对论。

爱因斯坦在引力问题上斗争的第一个焦点是潮汐引力。

[1] 评见书后第1章注释3。

[2] 牛顿的引力定律违反爱因斯坦的相对性原理，并不完全是显而易见的。因为爱因斯坦在建立这个原理时，依赖的是惯性参照系的概念，而这个概念不能用于引力存在的情况。（没有什么办法可以让一个参照系躲避引力而完全在自己的惯性影响下运动。）不过，爱因斯坦相信，一定有办法把他的相对性原理的影响扩大到引力的领域（也就是说，有某种办法将它“推广”从而将引力效应囊括进来），并且他还相信，牛顿引力定律将违反这个尚未建立的“推广的相对性原理”。

[3] 见第4章，特别是卡片4.1。

## 潮汐引力和时空曲率

想象你是一个正在遥远太空的宇航员，自由地向着地球赤道落下。尽管你在下落中感觉不出自己的重量，事实上你还是可以感觉某些小小的剩余的引力效应。这些剩余效应叫“潮汐引力”。我们先以地球上的某个观察者的观点，然后以你自己的观点来考虑你所感觉的引力。

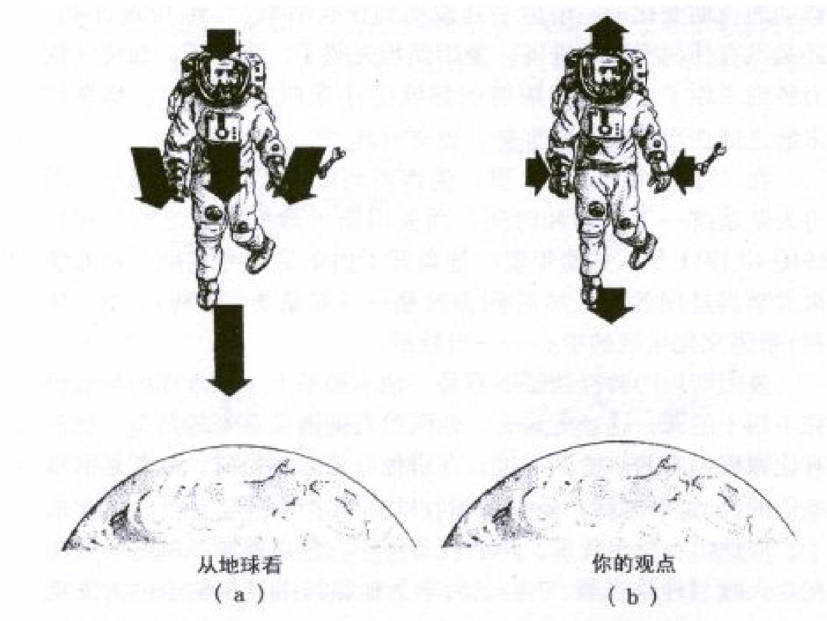



图2.3　当你落向地球时，潮汐引力会从头到脚将你拉长而又从两肋将你挤扁。

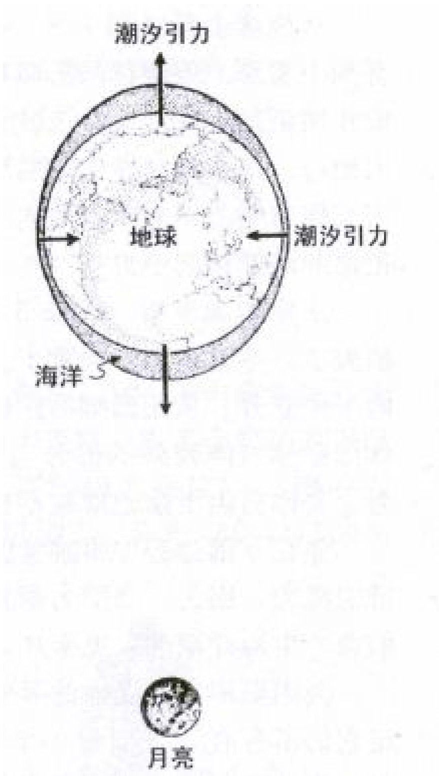
从地球上看（图2.3（a）），作用在你身上不同部位的引力有微小差别。因为你的脚离地球更近，引力对它们的作用比对头的作用更强，所以会从头到脚将你拉长。又因为引力作用总是指向地心，这个方向在你的右侧偏左，在你的左侧偏右，于是，作用在你右侧的引力有点儿向左，而左侧的向右，也就是说，引力把你的两侧挤向中央。

从你的观点看（图2.3（b）），巨大的向下的引力没有了，消失了。你觉得自己失重了。然而，消失的那部分引力只是拉你向下的部分，从头到脚的拉伸和两肋的挤压依然存在着，原因是作用在你身体较外的部分与作用在你身体中心的引力之间的差异，是你自由下落也摆脱不掉的。

你在下落过程中所感觉的垂向拉伸和侧向挤压，叫潮汐引力或引潮力，因为，当引力源是月球而让地球代替你来感觉时，它们会产生海洋潮汐，见卡片2.5。

爱因斯坦在演绎他的等效原理时，没有考虑潮汐引力，他假定它们不存在。（回想一下他论证的基本内容：当你自由下落时，你“不仅感觉不到自己的重量”，而且“你还会在所有方面都感到，似乎引力完全从你的邻近消失了”。）爱因斯坦忽略潮汐引力之所以是正当的，是因为他想象你（和你的参照系）很小。例如，假如你像蚂蚁那么小或者更小，那么你身体的各部分会彼此靠得很近，从而作用在身体外部和中心的引力方向几乎是一样的，引起潮汐拉伸和挤压的引力差异会极端地微弱。反过来讲，如果你是5 000公里高的巨人，那么地球作用在你身体外部和中心的引力在方向和强度上都将产生巨大的差异，当你下落时，你会经受剧烈的潮汐拉伸和挤压。

根据这样的推理，爱因斯坦相信，在自由下落的足够小的参照系（与引力作用变化的范围相比很小的参照系）中，我们不可能探测到任何潮汐引力的影响，也就是说，在我们的引力宇宙中，自由下落的小参照系与无引力宇宙中的惯性系是等效的，但对大参照系就不是这样了。而大参照系所感觉的潮汐力对1911年的爱因斯坦来说，似乎是最终认识引力本质的一个关键。




卡片2.5　潮汐力产生的海洋潮汐

在地球离月球最近的一边，月球的引力比作用在地心的更强，所以它比对固体地球更强烈地将海洋拉向月球，而海洋也会涌向月球。在地球离月球最远的一边，月球引力较弱，所以它对海洋的吸引不如对固体地球那么强烈，海洋也会凸出而远离月球。在地球的左侧，指向月心的引力有一向右的小分量，而在右侧，它有向左的分量，这些分量将海洋向内挤压。当地球自转时，海洋因为这个凸起和挤压的模式，在每天产生两个高潮和低潮。

也许，在你喜欢的海滨，潮汐并不完全是这样活动的，这不是月球引力的失败，而可能因为如下的两个效应：（1）海水对潮汐引力的反应有一定滞后，它在海湾、港口、河道、狭湾等沿海岸线的缺口

流进流出，需要时间；（2）太阳引力对地球的拉伸和挤压作用与月球几乎是一样强的，但是因为太阳在空中的位置（通常）与月球不同，所以引力作用的方向不同。地球潮汐是太阳和月球的潮汐引力联合作用的结果。

牛顿引力定律怎样解释潮汐力，现在清楚了：它们是作用在不同地方的引力在强度和方向上产生差异的结果。但牛顿的定律却因为它的引力依赖于距离而必将是错的，它违反了相对性原理（“这个距离应在谁的参照系中测量呢？”）。爱因斯坦的挑战是建立一个全新的引力定律，它可以同时满足相对性原理并以一种简单而令人信服的新方法来解释潮汐引力。

从1911年中期到1912年中期，爱因斯坦试图通过假设时间卷曲而空间平直来解释潮汐引力。这个听起来很极端的想法是引力时间膨胀的自然产物；天花板附近与地板附近时间流的不同速率可以想象为时间的卷曲。爱因斯坦猜测，也许更复杂的时间卷曲模式能产生从潮汐引力到行星椭圆轨道甚至水星近日点反常移动的所有已知的引力效应。

在追寻这个有趣的想法12个月后，爱因斯坦把它放弃了，当然他有很好的理由。时间是相对的，你的时间是我的时间和空间的混合（假如我们彼此相对运动），于是，如果你的时间是卷曲的而你的空间是平直的，那么我的时间和空间将都是卷曲的，其他任何人的也一样。你而且只有你才有平直的空间，所以物理学定律一定会将你的这个与其他参照系根本不同的参照系驱逐出去——因为它违反了相对性原理。

不过，凭爱因斯坦的感觉，时间卷曲“味道不错”，那么，也许——他想——每个人的时间都是卷曲的，相应的不可避免的是，每个人的空间也是卷曲的。也许这样联合的卷曲可以解释潮汐引力。

时间和空间两个都卷曲的想法是很吓人的。因为宇宙允许有无穷多个不同的参照系，每一个都以不同速度运动，那么将不得不有无穷多个卷曲的时间和无穷多个卷曲的空间！幸运的是，爱因斯坦认识到，闵可夫斯基已经为简化这个复杂的状态提供了有力的工具：“从今往后，空间和时间本身都将注定在黑暗中消失，只有二者的一种结合能保持为一个独立的实体。”在我们的宇宙中，只有一个惟一的绝对的四维时空，而每个人的时间和空间的卷曲，必然表现为闵可夫斯基单独的惟一的绝对时空的一种卷曲。

这是爱因斯坦在1912年夏天被迫得到的结论（不过他更喜欢说“曲率”，不说“卷曲”）。4年来，他一直在嘲笑闵可夫斯基的绝对时空，最后，他终于被迫接受了它，并让它发生卷曲。

时空会弯曲（或卷曲），是什么意思？为了讲清楚，我们先问，二维面的弯曲（或卷曲）意味着什么？图2.4画了一个平面和一个曲面。在平面（一张普通的纸）上画了两条绝对直的线，两条线并列延伸，是平行的。古希腊数学家欧几里得（他创立了现在称为“欧几里得几何”的学科）曾将两条初始平行的直线永不相交的要求作为他的一个几何假设。对平行直线所在的面来说，永不相交是确认面的平直性的铁证。如果空间是平直的，那么初始平行的直线永远不会相交。如果我们找到一对原先平行的直线确实相交了，那么我们将知道，空间不是平直的。

图2.4中的曲面是地球的球面。我们在球面上找到厄瓜多尔首都基多，它坐落在赤道上。从基多出发，画一条指向北方的完全直线，直线将在同一经度上向北延伸，穿过北极。

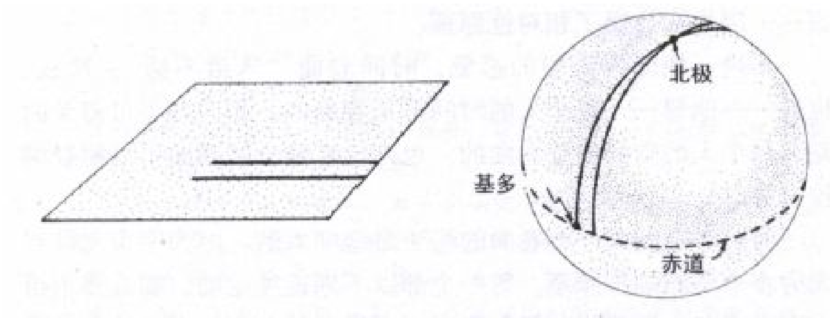


图2.4　在像左图的纸片那样的平面上，两条原先平行的直线永不相交。在像右图的地球球面那样的曲面上，两条原先平行的直线通常会相交。

在什么意义上说它是一条直线呢？有两种意义。一种意义对航空公司来说是极其重要的：直线是一个大圆，而地球球面上的大圆是两点间的最短路线，也就是航空公司愿意飞行的路线。任意另画一条联结基多与北极的路线，一定会比大圆长。

第二种平直性的意义我们在下面讨论时空时还会用到：在球面上沿大圆路径足够小的区域内，球面的曲率几乎测不出来。在这个区域，大圆看来是直的，就像我们通常在平坦的纸上所说的直线那样——这也是专业测量员在用经纬仪或激光束确定地产边界时所用的直线意义。在测量员的这个意义上，在沿大圆路径的每个区域内，大圆都是直线。

在弯曲或卷曲面上的任何路径，如果在这两种意义（航空上“最短路径”的意义和测量员的意义）上是直线，数学家就称它们是测地线。

现在，让我们在球面上从基多出发向东移动几厘米，画一条在赤道上完全与通过基多的那条线平行的新直线（大圆，测地线）。这条直线跟第一条一样，将经过球的北极。令这两条原先平行的直线后来在北极点相交的，正是球面的曲率。

明白了二维面上的曲率效应，我们就可以转向四维时空，去看看

那里的曲率。

在理想化的无引力宇宙中，既没有空间的卷曲，也没有时间的卷曲，时空没有曲率。根据爱因斯坦的狭义相对论，在这个宇宙中自由下落的粒子必然沿绝对的直线运动，在任何一个惯性参照系看来，它们都必然保持相同的方向和相同的速度。这是狭义相对论的基本原则。

现在，爱因斯坦的等效原理保证，引力不会改变自由运动的这一基本原则：当在我们真实的引力宇宙中自由运动的粒子进入并穿过一个小惯性（自由下落的）参照系时，它必然沿直线穿过参照系。然而，穿过小参照系的直线运动，显然类似于测量员在地球表面的一个小区域内所观测的直线行为；正如这种在地球小区域内的直线意味着直线实际上是地球表面的测地线一样，粒子在时空小区域内的直线运动也意味着粒子沿时空中的测地线运动。而这一个粒子经历的事情，对所有粒子也一定是正确的：每个自由运动的粒子（每个不受引力之外的任何力作用的粒子），沿时空测地线运动。

认识到这一点后不久，对爱因斯坦来说，潮汐引力是时空曲率的一个表现，就成为显然的事实了。

为说明这是为什么，我们来看下面的（我的，不是爱因斯坦的）思想实验。你一只手拿一个小球站在北极的冰层上（图2.5），同时将两球抛向空中，使它们沿精确的平行轨道上升，然后观察它们落回地球。现在，在我们这个思想实验中，你可以做你愿意做的任何事情，只要它不违反物理学定律。你不但想观察引力作用下的球在地球表面以上的轨迹，还想观察它在地表下的轨迹。为此，你可以假想球是特殊材料制成的，可以毫不减速地穿过地球的土壤和岩石（小黑洞可能具有这种性质），你还可以假想，你和一个站在地球另一端观察的朋友，可以通过“X射线图像”跟踪球在地球内部的运动。

球落入地球后，会因地球的潮汐引力而挤到一起，就像下落的宇航员两肋被挤压一样（图2.3）。潮汐引力的强度正好使两球几乎精确地落向地心，并在那儿相撞。

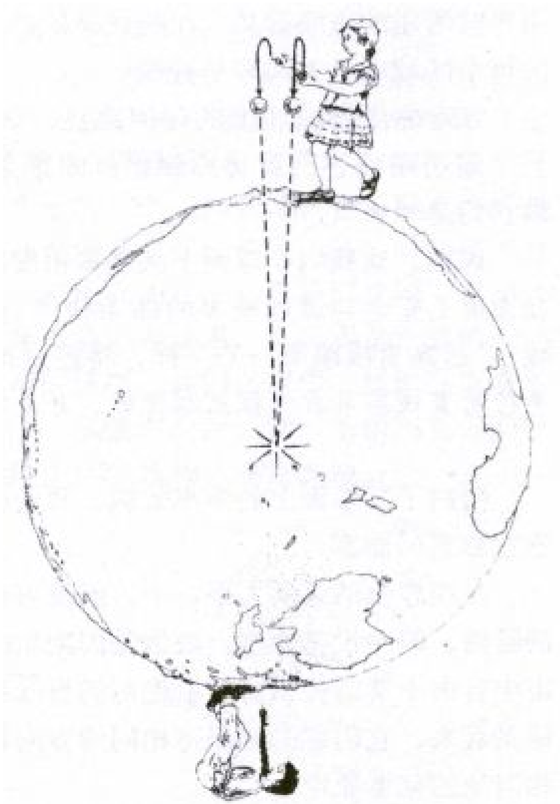


图2.5　两个沿精确平行路线抛入空中的球，如果能无阻碍地穿过地球，则它们会在地心附近相撞。

现在我们来总结一下这个思想实验：每个球在时空中沿完全的直线（测地线）运动，两条直线初始是平行的，后来相交了（球发生碰撞）。原先平行的直线相交了，这是时空曲率的标志。从爱因斯坦的观点看，时空曲率导致平行线相交，即导致两球相掩，就像图2.4中地球的曲率导致直线相交一样。从牛顿的观点符，潮汐引力导致两球相撞。

这样，因为在空间和时间本性上存在迥然不同的观点，爱因斯坦和牛顿对导致平行线相交的原因有完全不同的说法，爱因斯坦说它是时空曲率，牛顿说它是潮汐引力。但只有一个原因在起作用，因此，时空曲率和潮汐引力必然完全是以不同语言表达的同一件事情。

我们人类的头脑很难想象高于二维的曲面的图像，于是，几乎不可能形象地表现四维时空的曲率。不过，从不同的二维时空碎片，我们还是能看出一些事情。图2.6用两个时空碎片来解释时空曲率如何产生引起海洋潮汐的潮汐拉伸和挤压。

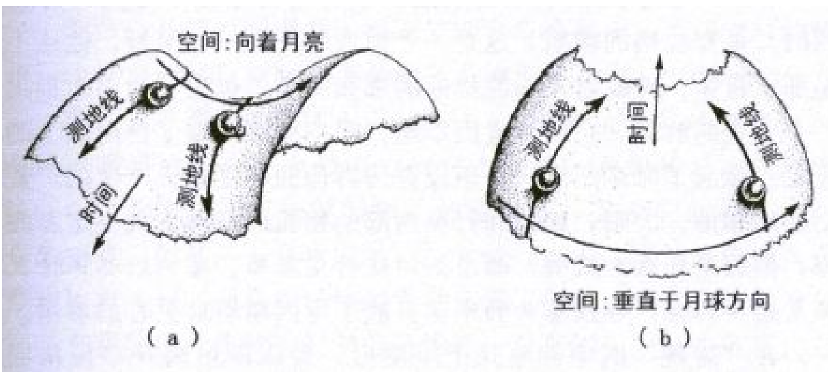



图2.6　地球附近的两片二维弯曲时空，曲率由月球产生。曲率在朝着月球的方向产生潮汐拉伸（a），在横向上产生挤压（b），拉伸和挤压以卡片2.5讨论的方式产生海洋潮汐。

图2.6（a）描绘的是地球附近的时空碎片，包括时间和朝月球方向的空间。月球使这块时空弯曲，曲率以图中所示的方式将两条测地线拉开。相应地，我们看到两个沿测地线旅行的自由运动粒子被拉开了，我们将这个拉开的作用解释为潮汐引力。拉伸作用的潮汐力（时空曲率）不仅影响自由运动粒子，也影响地球海洋，它像我们在卡片2.5中所看到的那样，在地球离月球最近和最远的一端掀起浪潮。浪潮也试图沿着弯曲时空的测地线运动（图2.6（a）），从而也试图飞起来分开，但地球的引力（地球产生的时空曲率，没画在图中）不让它们飞起来，所以海洋只能在地球上汹涌。

图2.6（b）是地球附近的另一块时空碎片，包括时间和沿垂直于月球方向的空间。月球使这块时空发生弯曲，曲率像图中那样将测地线挤压在一起。相应地，我们看到两个沿着垂直于月球方向的测地线旅行的自由运动粒子被曲率（月球的潮汐引力）挤在一起，同样，我们也看到地球海洋在垂直于月球方向的方向上被挤扁了。这种潮汐的挤压作用导致我们在卡片2.5中所看到的海洋的横向压缩。

1912年夏，爱因斯坦发现潮汐引力与时空曲率是同一样东西时，是布拉格的教授。这是一个惊人的发现——尽管，他还不是那么肯定，理解也不像我描述的那么完全，也没有为引力提出一个完全的解释。它告诉爱因斯坦，时空曲率决定了自由粒子的运动，掀起了海洋的潮汐，但没有告诉他曲率是怎么产生的。爱因斯坦相信，太阳、地球和行星内部的物质以某种方式决定着曲率，但那是什么方式呢？物质如何使时空卷曲，卷曲的具体情况又是怎样的呢？寻找卷曲的定律，成了爱因斯坦最关心的事情。

在“发现”时空曲率几个星期后，爱因斯坦离开布拉格回到苏黎世，到他的母校ETH当教授。1912年8月，爱因斯坦刚到苏黎世就去请教老同学格罗斯曼（Marcel　Grossmarnn），现在是那儿的数学教授。爱因斯坦向他解释了潮汐引力是时空曲率的思想，然后问他，有没有什么数学家已经建立的数学方程能帮他发现卷曲的定律，也就是描述物质如何令时空弯曲的定律。格罗斯曼没把握，他的专业在几何的其他方面。不过，去图书馆浏览后，他回来说，有的，确实有需要的方程。这些方程主要是德国数学家黎曼（Bernhard　Riemann）在19世纪60年代、意大利的里奇（Gregorio　Ricci）在80年代以及里奇的学生勒维-契维塔（Tullio　Levi-Civita）在19世纪90年代和20世纪初建立起来的，叫“绝对微分计算”（或者，用1915～1960年间物理学家的语言说，“张量分析”）。不过，格罗斯曼告诉爱因斯坦，微分几何太乱了，物理学家不应该卷进来。还有别的可以用来揭示卷曲定律的几何吗？没有。

就这样，在格罗斯曼的大力协助下，爱因斯坦决定掌握复杂的微分几何，格罗斯曼教爱因斯坦数学，爱因斯坦也教格罗斯曼一些物理学的东西。后来，爱因斯坦引格罗斯曼的话说，“我承认，我毕竟还是通过物理学习得到了一些相当重要的东西。以前，当我坐在椅子上感觉到‘前坐者’的余温时，总有点儿不舒服，现在这种感觉完全没有了。因为，在这一点上物理学家已经告诉了我，热是完全与个人无关的东西。”

学微分几何对爱因斯坦并不是件容易的事情。这门学科的精神同他那自然的直觉的物理学论证是格格不入的。1912年10月底，他给索末菲（Arnold Sommerfeld，一个德国大物理学家）写信说，“我自己现在完全被引力问题占据了，我也相信，在当地的一个数学家朋友[格罗斯曼]的帮助下，我会有能力克服所有的困难。但有一点是肯定的，在我的一生中，还从来没有这么艰难地奋斗过，而且我已经对数学充满了敬佩，它那精妙的部分至今在我简单的头脑中还只能认为是一种奢望！同这个问题比起来，原先的相对性理论[狭义相对论]不过是儿童游戏。”

面对物质如何令时空弯曲的疑惑，爱因斯坦同格罗斯曼一道，从秋天奋斗到冬天，但不论多大的努力，都没能使数学与爱因斯坦的想象相吻合，卷曲的定律在躲着他们。

爱因斯坦确信，卷曲的定律应该服从推广（扩大）形式的相对性原理：对每一个参照系——不仅是惯性（自由下落的）系，还包括非惯性系，它都应该是一样的。卷曲的定律不应为了自己的形式而依赖于任何特殊的参照系或者什么特殊类型的参照系。[1]痛苦的是，微分几何的方程似乎不允许有这样的定律。在暮冬时候，爱因斯坦和格罗斯曼最终放弃了寻找，发表了他们所能发现的最好的卷曲定律——为了某种确定性而依赖于一类特殊参照系的定律。

永远乐观的爱因斯坦努力使自己相信，简单说，就是相信这不是

什么灾难。1913年初，他在给朋友物理学家埃伦费斯特（Paul Ehrenfest）的信中说，“还有什么能比从[能量和动量守恒的数学方程]得到的必然限制更美妙的呢？”但进一步考虑后，他认为这是一个灾难。1913年8月，他写信给洛伦兹说，“我对这个理论[‘卷曲的定律’]的可靠性的信心还在动摇……[因为不能遵从一般的相对性原理，]这个理论背离了它自己的出发点，一切都悬而未决。”

当爱因斯坦和格罗斯曼在跟时空曲率斗争时，遍布欧洲大陆的其他物理学家担起了统一引力定律和狭义相对论的挑战。但是，他们——芬兰赫尔辛基的诺德斯特勒姆（Gunnar Nordstrφm），德国格赖夫斯瓦尔德的米（Gustar Mie），意大利米兰的亚伯拉罕（Max Abraham）——没有一个人采纳了爱因斯坦的时空曲率观点，而是像电磁那样将引力当成一种活动在闵可夫斯基的平直的狭义相对论时空中的力场。他们采用这种方法也并不奇怪：爱因斯坦和格罗斯曼所用的数学吓人地复杂，而且得到的卷曲定律违反了作者自己的原则。

在不同观点的拥有者之间，发生了激烈的论战。亚伯拉罕写道：“像本作者一样曾不得不反复告诫要警惕[相对性原理]的诱惑的人，将满意地欢迎这样的事实：它的创始者现在已经令自己相信它是不可能维持下去的。”爱因斯坦在答复中写道：“照我的意见，这种情况并不说明相对性原理失败了……没有一点儿根据怀疑它的有效性。”私下里，他把亚伯拉罕的引力理论描述成“一匹缺了三条腿的高贵马儿”。在1913年和1914年间写给朋友们的信中，爱因斯坦谈了这场争论，“我喜欢这个，至少事情有了必要的生气，我欣赏这种争论，费加罗式的：他想跳舞，我愿为他伴奏。[2]”“我很高兴，同行们都投身到这个[格罗斯曼和我发展起来的]理论中来了，尽管他们现在的目的是要抹煞它……从表面看，诺德斯特勒姆的理论……似乎合理得多，但它也是建立在[平直的闵可夫斯基时空]的，我感到，人们对[平直时空]的信仰，差不多像某种迷信了。”

1914年4月，爱因斯坦离开苏黎世，到柏林做一名不用讲课的教

授。他终于可以如愿地尽情工作了，甚至在柏林的大物理学家普朗克和能斯特（Walther Nernst）的影响里，他还是这么做。尽管在1914年爆发了第一次世界大战，爱因斯坦在柏林仍然继续追寻着一个能让人接受的关于物质如何使时空发生弯曲的描述，一个不依赖于任何特殊类型的参照系的描述——一个改进的关于卷曲的定律。

从柏林坐3个小时火车就来到闵可夫斯基曾经工作过的哥廷根大学村，历史上最伟大的数学家之一，大卫·希尔伯特（David Hilbert）就在那里。在1914和1915年间，他对物理学产生了强烈的兴趣。爱因斯坦发表的思想令他着迷。于是，1915年6月底，他邀请爱因斯坦来访问。爱因斯坦去了约一个星期，为希尔伯特和他的同事们作了6次两个小时的演讲。访问过了几天以后，爱因斯坦给朋友写信说，“看到[关于我工作的]每件事情在哥廷根都能得到最彻底的理解，我真太高兴了。希尔伯特也令我着迷。”

回柏林几个月后，爱因斯坦-格罗斯曼卷曲定律令他比以前更痛苦。它不但违反了他的引力定律应在所有参照系相同的理想，而且，经过艰难计算后，他还发现，得到的水星轨道近日点的异常移动值是错的。他原希望这个理论能够解释这个近日点移动，从而胜利解决移动与牛顿定律的偏差，这个成果至少能带来一点实验证据，证明他的引力定律是对的，而牛顿的是错的。然而，他在爱因斯坦-格罗斯曼卷曲定律基础上的计算却只得到了观察到的近日点移动的一半。

爱因斯坦彻底检查了他和格罗斯曼过去的计算，发现几个关键性的错误。整个10月，他都在满怀激情地工作。11月4日，他在柏林的普鲁士科学院周末全体会议上提交了关于他的错误和修正的卷曲定律的报告——定律对一类特殊的参照系仍存在一定依赖性，不过不像以前那么强。

爱因斯坦还是不满意，他又同11月4日的定律斗争了一个星期，发现了错误，向11月11日的科学院大会又提出一个卷曲定律的建议。

但是，定律还是依赖于特殊参照系，仍然违背他的相对性原理。

就让它违背下去吧。接下来的一个星期里，爱因斯坦计算了他的新定律的可以通过望远镜观测的结果。他发现定律预言，经过太阳边缘的星光应被引力偏转1.7弧秒的角度（4年以后在一次日食中进行的精确测量将证实这个预言）。而对爱因斯坦最重要的是，新定律给出了正确的水星近日点移动！他欣喜若狂，兴奋得3天做不了事情。在11月18日的科学院大会上，他报告了这个胜利。

但是，他的定律违背相对性原理，这仍令他烦恼。于是，他在下个星期又检查了计算，发现了另外的错误——最关键的一个。终于，一切都明白了，整个数学体系现在都摆脱了对特殊参照系的任何依赖：在任意一个参照系中定律都有相同的形式（见下面的卡片2.6），因此服从相对性原理。爱因斯坦1914年的理想完全实现了！新的公式对水星近日点进动和光的引力偏折给出相同的预言，而且把他1907年的引力时间膨胀的预言也包括进来了。11月25日，爱因斯坦向普鲁士科学院报告了他这些结果和他的广义相对论的最终确定形式。[7]

3天后，爱因斯坦给朋友索末菲写信说：“在过去的一个月里，我度过了一生中最兴奋、最艰苦但也最成功的时光。”接着，在[1916年]1月给埃伦费斯特的信中，他说：“你能想象我有多快乐，[我的新的卷曲定律遵从相对性原理，]它还预言了正确的水星近日点的运动。我狂喜了几天。”后来，他谈了这个时期的感受：“在黑暗中找寻我们感觉得到却表达不出的真理的年月里，那强烈的欲望和动摇的信心以及成功前的焦虑，只有亲身经历过的人才能体会。”

值得注意的是，爱因斯坦并不是第一个发现卷曲定律的正确形式（服从相对性原理的形式）的人，第一个发现者应该是希尔伯特。1915年秋，当爱因斯坦还在向正确定律努力，数学错误接连不断时，希尔伯特也在考虑他从爱因斯坦夏季来哥廷根访问时学来的东西。他

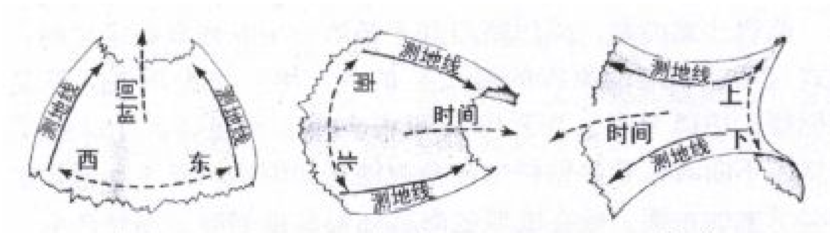
在波罗的海的陆根岛度假时，突然产生一个关键的想法，几个星期后，他得到了正确的定律——他没有走爱因斯坦那条艰难的试错路线，而是走一条优美而简捷的数学道路。1915年11月20日，希尔伯特向哥廷根的皇家科学院报告了他的推导和最后的定律，正好比爱因斯坦在柏林向普鲁士科学院报告相同定律早5天。

不过，最后的卷曲定律很快被称为爱因斯坦场方程（卡片2.6），而没有用希尔伯特的名字来命名，这是很自然的，也符合希尔伯特自己对事情的看法。希尔伯特独立并几乎与爱因斯坦同时完成了这个发现的最后几个数学步骤，但爱因斯坦发现了这些步骤之前的几乎一切东西：认识了潮汐引力与时空卷曲必须是同一件事情，设想卷曲定律必然服从相对性原理，它们是这个定律（爱因斯坦场方程）的90%。事实上，如果没有爱因斯坦，引力的广义相对论定律可能会晚发现几十年。

卡片2.6　爱因斯坦场方程：爱因斯坦的时空卷曲定律[8]

爱因斯坦的时空卷曲定律，即爱因斯坦场方程指出，“物质和压力使时空卷曲。”更具体地说：

在时空任一位置任选一个参照系，通过研究在这个选定的参照系的三个方向（东-西、南-北和上-下）上曲率（即潮汐引力）将自由运动的粒子推近或拉开的方式来寻找时空的曲率。粒子沿时空的测地线运动（图2.6），它们被推近或拉开的速率正比于它们之间的方向上的曲率大小。如果它们像在图（a）和（b）中那样被推近，我们就说曲率是正的；如果它们像在图（c）中那样被拉开，曲率就是负的。




(a)　　(b)　　(c)

将东-西（图（a））、南-北（图（b））和上-下（图（c））三个方向上的曲率加起来，爱因斯坦场方程指出，这三个曲率大小的和正比于粒子附近的质量密度（乘以光速的平方化为能量密度，见卡片5.2）加上粒子附近物质压力的3倍。

即使你和我可能处在时空的同一位置（如，1996年7月14日中午飞在法国巴黎上空），如果我们彼此相对运动，你的空间将不同于我的，同样，你测量的质量密度（如我们周围空气的质量）也将不同于我测量的密度，我们测量的物质压力（如空气压力）也将不同。同样，你测量的三个时空曲率之和也将不同于我测量的和。然而，你和我都一定会发现，我们测量的曲率之和正比于我们测量的质量密度加上我们测量的压力的3倍。在这个意义上，爱因斯坦场方程在每个参照系中都是一样的；它服从爱因斯坦的相对性原理。

在大多数情况下（如整个太阳系），物质的压力与它的质量密度乘以光速平方栢比很小，因而压力对时空曲率的贡献是不重要的；时空卷曲几乎只归因于质量。只有在中子星内部深处（第3章）和其他特别的地方，压力对卷曲的贡献才有意义。

通过爱因斯坦场方程的数学计算，爱因斯坦和其他物理学家不但解释了光线被太阳的偏折和行星在轨道上的运动（包括奇怪的水星近日点的移动），而且还预言了黑洞（第3章）、引力波（第10章）、时空奇点（第13章）的存在，也许还有虫洞和时间机器（第14章）的存在。本书其余部分就用来讨论爱因斯坦这些天才的遗物。

当我浏览爱因斯坦发表的科学论文时（很遗憾，我只能看1965年的俄文本选集，因为我不懂德文，而他的大多数论文到1993年才开始译成英文！）[9]，我惊讶地发现他的研究风格在1912年发生了巨大的改变。1912年前，他的文章以优美的文笔、深刻的直觉和简单的数学而令人赏心悦目，其中许多论证跟我们在90年代讲相对论时所用的是一样的，没人想去改进它们。相反，1912年以后，爱因斯坦的论文里出现了大量复杂的数学——尽管通常结合着他对物理学定律的洞察。这种数学与物理学洞察的结合，在1912～1915年间所有在引力领域工作的物理学家中，只有爱因斯坦才有，它最终将他引向了引力定律的完全形式。

但是，爱因斯坦的数学工具用得有点儿笨拙，像希尔伯特后来说的，“哥廷根街上的每个小孩儿都比爱因斯坦更理解四维几何。不过，尽管如此，爱因斯坦还是做成了这件事[建立引力的广义相对论定律]，而数学家没有。”他之所以做成了，是因为这件事仅有数学是不够的；还需要他那独特的物理学洞察。

实际上，希尔伯特说得过分了。爱因斯坦是一个很不错的数学家，尽管，他在数学技巧上不像在物理学洞察中那样算得上是一个大师。结果，我们现在很少用爱因斯坦1912年以后提出的形式来讨论他当时的论证，我们已经学会了更好的形式。而且，随着1915年过后，物理学定律的数学味道越来越浓，爱因斯坦当年那个巨人的身影也越来越淡，火炬已经传给了别人。

[1] 爱因斯坦用一个新名词“广义协变性”来称这种性质，尽管它不过是他的相对性原理的自然推广。

[2] 这里，爱因斯坦引用了歌剧《费加罗的婚礼》（Lorenzo Da Ponte编剧，莫扎特曲，见第二幕）里的台词。——译者

# 第3章　黑洞，发现与拒绝

爱因斯坦的卷曲时空定律

预言了黑洞，

爱因斯坦拒绝了这个预言

“至于为什么‘史瓦西奇点’不存在于物理学实体中”，1939年，爱因斯坦在一篇论文中写道，“这个考察的基本结果说得很清楚了。”[1]爱因斯坦用这句话明确地拒绝了他自己的理性财产：他的广义相对论引力定律似乎正在预言的黑洞。

那时，根据爱因斯坦的定律还只能得到黑洞的几个性质，而“黑洞”这个名字也还没有，它们被称为“史瓦西奇点”。不过，人们已经明白，落入黑洞的任何事物不可能再逃出来，也不可能发出光或其他东西，而这已经足以让爱因斯坦和他那时代的大多数物理学家相信，黑洞是可怕的怪物，肯定不会存在于真实宇宙中。物理学定律一定会以某种方法使宇宙不受这种怪物的侵害。

爱因斯坦如此强烈地拒绝黑洞，那么，关于黑洞，他们那时都知道些什么呢？广义相对论关于黑洞存在的预言有多大力量？爱因斯坦怎么能拒绝这个预言而仍然相信他的广义相对论的定律呢？这些问题的答案有着18世纪的渊源。

在整个18世纪，科学家（那时叫自然哲学家）们相信，引力服从牛顿定律，光由光源以极高的普适速度发出的微粒（粒子）组成。通过望远镜对木卫在绕木星的轨道上所发出的光的测量，知道光速大约

是每秒300 000千米。

1783年，英国自然哲学家米歇尔（John Michell）大胆地将光的微粒描述与牛顿的引力定律结合，从而预言了非常致密的星体应该是什么样的。[2]我把他的思想实验换个说法重复一遍：

在一颗星体的表面以某初始速度抛出一个粒子让它自由向上运动。如果初始速度太低，星体引力将减慢粒子速度，使它停下来，然后将它拉回星体表面。如果初始速度足够大，引力也将使粒子慢下来，但不会使它停止，粒子将设法逃掉。回落与逃逸的界线，即为了逃逸的最小初始速度，叫“逃逸速度”。对从地球表面抛出的粒子来说，逃逸速度是每秒11千米，从太阳表面抛出的粒子，逃逸速度为每秒617千米，或光速的0.2%。

米歇尔能用牛顿的引力定律计算逃逸速度，证明它正比于星体质量除以其周长的平方根。因此，对质量一定的星体来说，周长越小，逃逸速度越大。理由很简单：周长越小，星体表面离中心越近，因而表面的引力越强，粒子为了逃脱星体的引力作用就越困难。

米歇尔推论，存在一个临界周长，对它来说逃逸速度是光速。如果光微粒像其他类型的粒子一样受引力作用，那么光几乎不能从具有临界周长的星体逃逸出去。对更小的星体光就完全不能逃逸了。如果以标准光速299 792千米/秒从这样的星体发射一颗光微粒，微粒起初会向上，然后慢慢停下来，又落回星体表面，见图3.1。

米歇尔能够很容易地计算临界周长。假如星体与太阳有相同质量，那么周长是18.5千米，而且随质量成比例地增大。




图3.1　米歇尔1783年用牛顿引力定律和光的微粒描述计算的从比临界周长小的星体发出的光的行为。

18世纪的物理学定律无法阻止如此致密星体的存在，因此，米歇尔猜想，宇宙中可能存在大量这样的黑（暗）星，它们都圆满地存在于自己的临界周长内，从地球看不到它们，因为从它表面发出的光微粒都被无情地拉回去了。这样的暗星就是黑洞在18世纪的形式。

米歇尔是英格兰约克郡桑希尔的教区长，1783年11月27日，他向皇家学会报告了也许存在暗星的预言。报告在英国自然哲学界产生了一点影响。13年后，法国自然哲学家拉普拉斯（Pierre Simon Laplace）在他的名著《宇宙体系论》的第1版里，通俗地提出了相同的预言，而没有提到比他更早的米歇尔的工作。在1799年的第2版里，拉普拉斯还保留了暗星的预言；但是到第3版（1808年）的时候，托马斯·杨（Thomas Young）发现了光的自干涉，[1]证实了惠更斯（Christiann Huygens）提出的光的波动描述，从而迫使自然哲学家们放弃光的微粒描述——不过那时还不知道如何协调波动描述与牛顿引力定律，以计算星体引力对它发出的光的作用。大概为了这个原因，拉普拉斯在第3版和后来的版本中删除了暗星的概念。[3]

只有在1915年11月爱因斯坦建立了广义相对论的引力定律以后，物理学家才又一次相信他们对引力和光的认识已经足以计算星体引力对它发出的光的作用了。只有在这个时候，他们才又满怀信心地

转到米歇尔和拉普拉斯的暗星（黑洞）上来。

第一步是史瓦西（Karl Schwarzschild）迈出的，他是20世纪初最有名的天体物理学家。那时，他正在第一次世界大战俄国前线的德国军队服役，在1915年的《普鲁士科学院会议报告》里看到了爱因斯坦建立的广义相对论。他几乎立即就开始寻找爱因斯坦的新引力定律对星体能作出什么预言。

由于分析旋转的或非球形的星体在数学上很复杂，为了简化计算，史瓦西只考虑了完全没有旋转的球状星体，他先去找星体外部的数学描述，然后再来揭示星体的内部。几天之内，他就找到了答案。他根据爱因斯坦的新场方程，详细计算了任意无旋转球状星体外的时空曲率。他的计算简洁而优美，计算所预言的弯曲几何，很快成为大家所熟悉的史瓦西几何，注定会对我们认识引力和宇宙产生巨大的影响。

史瓦西把计算论文寄给爱因斯坦。1916年1月13日，爱因斯坦代表他在柏林普鲁士科学院的一次会议上作了报告。几个星期后，爱因斯坦向科学院报告了史瓦西的第二篇论文：关于星体内部的时空弯曲的精确计算。[4]仅仅4个月后，史瓦西令人瞩目的成果终止了：6月19日，爱因斯坦悲痛地向科学院报告，卡尔·史瓦西在俄国前线染病去世了。

史瓦西几何是我们在本书遇到的第一个时空弯曲的具体例子，而且对黑洞性质非常重要，所以我们要详细地来考察它。

假如我们把在空间和时间各处的所有活动都想象为一个绝对的统一的四维时空“结构”，那么，用弯曲（卷曲）的四维时空语言来描述史瓦西几何真是再恰当不过了。然而，我们日常经历的是三维空间与一维时间，它们是没有统一的；所以，我在描述中，将卷曲的时空分解为卷曲的空间加卷曲的时间。

因为空间和时间是“相对的”（如果我们彼此相对运动，则我的空间不同于你的空间，我的时间也不同于你的时间2），时空的分解首先要求选择一个参照系——也就是选择一种运动状态。对某颗星而言，我们有一种自然的选择，即令星体处于静止，也就是，我们选择星体自己的参照系。换句话说，我们来检验星体自己的空间和时间而不是某个高速经过星体的运动者的空间和时间，这是很自然的。



身穿学者长袍的卡尔·史瓦西在德国哥廷根。

[美国物理学联合会（AIP）Emilio Segrè图像档案馆提供。]

为了更形象地认识星体空间的曲率（卷曲），我将借助一种叫嵌

入图的绘图方法。因为嵌入图将在以后的章节里充当重要角色，我现在通过类比来仔细介绍这个概念。

想象一组生活在只有两个空间维的宇宙中的类人生物，他们的宇宙是一个弯曲的碗面，如图3.2所示；他们也跟他们的宇宙一样，是二维的，在垂直于曲面的方向上，是无限薄的。而且，他们看不到曲面的外头，他们通过在曲面上运动而永不离开曲面的光来观察事物。于是，这些“二维生物”（我这么称呼他们）没有任何方法来获得关于他们二维宇宙以外的任何事物的任何信息。

二维生物可以通过对直线、三角形和圆的测量来探索他们二维宇宙的几何。他们的直线是第2章讨论过的“测地线”（图2.4及有关正文）：他们二维宇宙中存在的最直的线。在他们宇宙的“碗”底（在图3.2中我们看到是球面的一部分），他们的直线是像地球赤道或者经线一样的大圆的一部分。在碗口以外，他们的宇宙是平坦的，所以那里的直线是我们通常所认识的直线。

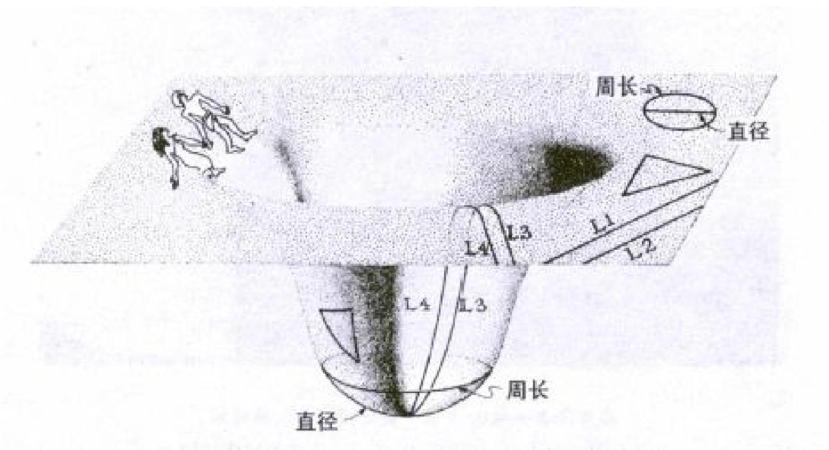



图3.2　居住着二维生命的二维宇宙

如果二维生物去检验他们宇宙外部平直部分的一对平行直线（如图3.2中的L1和12），那么不论他们跟踪直线多远，都不会看到它们

相交。用这个办法，他们发现外部区域是平直的。另一方面，如果他们在碗口外作平行直线L3和L4，然后跟踪两线进入碗内，尽可能让它们保持为直线（测地线），他们将看到线在碗底相交。他们用这个办法发现，在碗的内部区域，他们的宇宙是弯曲的。

通过测量圆和三角形（图3.2），二维生物也可以发现外部区域是平直的，而内部区域是弯曲的。在外部区域，所有圆的周长都等于（3.141 592 65……）乘以它们的直径。在内部区域，圆的周长小于π乘以圆的直径。例如，图3.2中近碗底的大圆的周长等于2.5乘以直径。如果二维生物以直线（测地线）为边作一个三角形，然后将三个内角相加，那么，在平直的外部区域，结果是180度；而在弯曲的内部区域，结果会大于180度。

通过这些测量，二维生物发现了他们的宇宙是弯曲的，接着他们开始推测，可能存在一个三维空间，他们的宇宙就处在——即嵌在其中。他们可能将那个三维空间叫超空间并猜想它的性质。例如，他们可能想象它在欧几里得意义上是“平直的”（其中的平行线永不相交）。你我看这个空间都没有困难，它就是图3.2中的三维空间，我们每天经历的空间。然而，二维生物因为只有二维经验，他们来看这个空间是大有困难的。而且，他们不可能有什么办法知道，这样的超曲面是否真的存在。他们不可能走出二维宇宙进入超曲面的第三维，又因为只能通过永远在宇宙中的光线来观察，所以他们永远也看不到超曲面。对他们来说，超曲面完全是假想的。

超曲面的第三维与二维生物可能也会认为是第三维的“时间维”无关。当他们想象超曲面时，实际不得不用四维的语言：两维是他们宇宙的空间，一维是它的时间，还有一维，就是超曲面的第三维。

我们是三维生物，生活在一个弯曲的三维空间里。如果我们也去测量我们在一个星体内部和附近的空间几何，史瓦西几何——我们将发现它会像二维生物的宇宙那样弯曲。

我们可以想象一个更高维的平直超曲面，我们弯曲的三维空间就嵌在其中。为了适应像我们这样的三维弯曲空间，结果，这样的超曲面必须有六个维。（记住，我们的宇宙也有一个时间维，所以总体上我们必须在七维下进行思考。）

现在，要我将我们嵌在六维超曲面中的三维空间形象地表现出来，就比二维生物表示他们嵌在三维超曲面中的二维宇宙，困难得多了。不过，有一种技巧可以帮很大的忙，如图3.3。

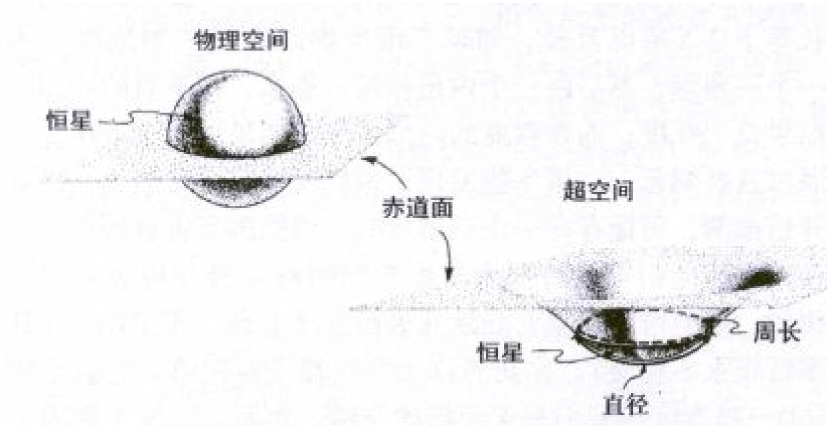



图3.3　用嵌入图（右下）表示的星体内部和周围（左上）的三维空间的曲率，这是爱因斯坦场方程的史瓦西解所预言的曲率。

图3.3描述了一个思想实验：将一张薄片插入一个星体的赤道平面（左上图），薄片将星体切成完全相同的上下两半。赤道薄片虽然在图中看起来还是平的，但实际上是不平的。星体的质量使星体内部和周围的三维空间发生卷曲（左上图不能表现卷曲方式），卷曲又以图中没有表现的方式使赤道片发生弯曲。我们可以完全像二维生物在他们宇宙的二维空间那样，在我们真实的物理空间中进行几何测量，从而发现薄片的曲率。这样的测量将揭示，原先平行的直线在星体中心附近相交，任何星体内部或附近的圆的周长小于π乘以其直径，三角形内角和大于180度。弯曲空间的这些古怪事情都是爱因斯坦方程

的史瓦西解预言的。

为了形象地表现史瓦西曲率，我们可以像二维生物那样，想象将赤道片从我们宇宙弯曲的三维空间中取出来，然后将它嵌入一个假想的平直的三维超曲面（图3.3右下）。在没有弯曲的超曲面中，薄片只有向下弯曲成碗状，才能保持它的弯曲几何。这样从我们弯曲宇宙中取出来嵌入假想的三维平直超曲面的二维薄片图，就叫嵌入图。

将超曲面的第三维想象为我们自己宇宙的第三个空间维是很诱人的，但我们必须拒绝这种诱惑。超曲面的第三维与我们自己宇宙的任何一维都没有任何关系，那一维我们既走不进也看不见，也不能从它得到任何信息，是纯假想的一维。不过，它还是有用的。它可以帮助我们看见史瓦西几何，而且在本书后面它还将帮助我们看见其他的弯曲几何：黑洞的、引力波的、奇点的和虫洞的（见6，7，10，13和14章）。

如从图3.3的嵌入图所看到的，星体赤道片的史瓦西几何在定性上同二维生物宇宙的几何是一样的：在星体内部，几何是碗状弯曲的，在远离星体的地方，几何是平直的。跟在二维生物的碗内的大圆（图3.2）一样，在这儿（图3.3），星体的周长除以直径也小于π。对我们的太阳而言，它预言周长与直径之比比π小百万分之几；换句话说，在太阳内部，空间在百万分之几以内的精度上是平直的。然而，假如让太阳保持相同的质量而周长越来越小，那么它内部的曲率会越来越大，图3.3的嵌入图的碗的下凹将更显著，周长与直径之比将大大地小于π。

因为在不同参照系中空间的曲率不同（“如果我们彼此相对运动，你的空间是我的空间和我的时间的混合”），所以在相对于星体以高速度运动的参照系中和在星体处于静止的参照系中所测量的星体空间弯曲的细节是不同的。在高速参照系的空间里，星体在垂直于运动的方向上以某种方式被压扁了，所以嵌入图很像图3.3的，但碗被

横向挤压成为长椭圆形。压扁是弯曲空间发生了菲兹杰拉德在无引力宇宙中所发现的空间收缩的结果（见第1章）。

爱因斯坦场方程的史瓦西解不仅描述了空间的曲率（或卷曲），还描述了星体附近的时间卷曲——由星体的强大引力产生的卷曲。在相对于星体静止或飞行速度不高的参照系中，时间卷曲完全是第2章讨论过的引力时间膨胀（卡片2.4及相关讨论）：靠近星体表面的时间流比远处的慢，而在星体中心，时间流更慢。

在太阳的情形，时间卷曲很小：在太阳表面，时间流只比远离太阳慢百万分之二（1年慢64秒），而在太阳中心，它比在远处慢大约十万分之一（1年慢5分钟）。然而，如果太阳保持相同质量而周长更小，表面离中心更近，那么引力作用会更强，相应地，它的引力时间膨胀——时间卷曲——将变得更大。

时间卷曲的一个结果是，从恒星发出的光会经历引力红移。因为光的振荡频率由光发射处的时间流决定，从星体表面的原子发出的光在到达地球时，将比从星际空间的同类原子发出的光具有更低的频率。频率降低的量完全与时间流变慢的数量相同。较低的频率意味着较大的波长，所以，来自星体的光必然会以星体表面时间膨胀的数量向光谱的红端移动。

在太阳表面，时间膨胀为百万分之二，所以，从太阳到达地球时，光的引力红移也是百万分之二。在爱因斯坦时代，这么小的红移是不能确定地测量的。但在60年代初，实验技术赶上了爱因斯坦的引力定律：普林斯顿大学的布劳特（Jim　Brault）用一个精巧的实验测量了太阳光的红移，得到了与爱因斯坦预言非常一致的结果。[5]

在史瓦西过早去世后的几年里，他的时空几何成了物理学家和天体物理学家的标准研究工具。包括爱因斯坦在内的大多数人都在研究它，估量它的意义。所有的人都同意而且重视这样的结论，如果星体

像太阳那样有很大的周长，那么它内部和周围的时空只会出现很小的弯曲，从它表面发出而在地球接收的光向红色方向的移动会更少。他们也同意，如果星体越致密，那么它的时空卷曲越大，从它表面发出的光的引力红移也越大。不过，很少有人愿意认真去考虑史瓦西几何在高致密星体情况下的那些极端预言（图3.4）。[6]

史瓦西几何预言，每个星体都存在一个依赖于星体质量的临界周长——与米歇尔和拉普拉斯在一个多世纪以前发现的那个临界周长相同：18.5公里乘以以太阳质量为单位的星体质量。如果星体的实际周长大于这个临界值的4倍（图3.4上），那么星体空间将像途中表现的那样适度弯曲，它表面的时间流会比远处的慢15%，从表面发出的光也会向红端移动15%。如果星体周长较小，只是临界值的2倍（图3.4中），它的表面将更强烈地弯曲，表面的时间会比远处的慢41%，表面发出的光也将红移41%。这些预言看来都是合理的，可以接受的。在20年代甚至60年代末的物理学家和天体物理学家们看来根本不合理的，是在实际周长与临界值相同的星体情况下的预言（图3.4下）。这样一个星体，因为它更强烈弯曲的空间，表面的时间流无限地膨胀，时间根本不流了——冻结了。相应的是，不论从星体表面出发向上旅行的光开始是什么颜色，它都一定会移过光谱的红端，超越红外线，超越无线电波的波长，一直到无限大，就是说，直到不存在。用现代语言来说，具有临界周长的星体表面正好处在黑洞的视界；而星体以它强大的引力产生一个包围它自身的黑洞视界。

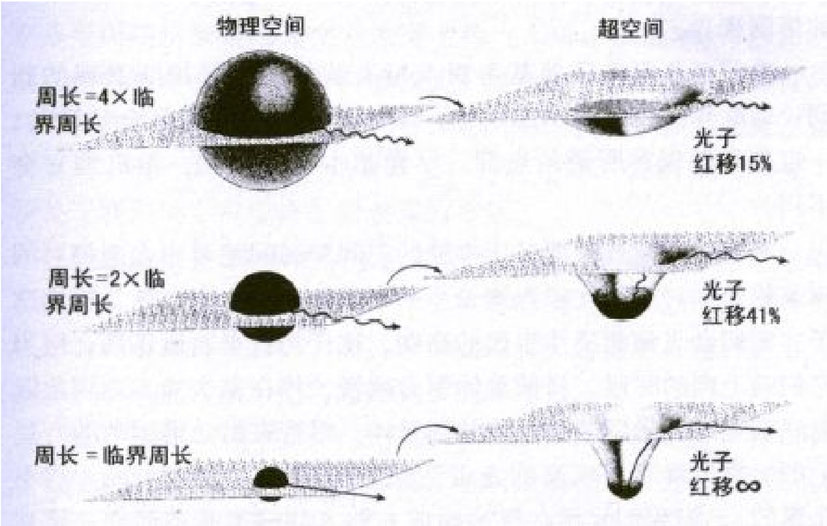



图3.4　广义相对论对空间曲率和光的红移的预言，光线来自3个质量相同而周长不同的高度致密的星体。第1个星体的周长是临界周长的4倍；第2个是临界周长的2倍；第3个正好具有与临界值相同的周长。用现代语言说，第3个星体的表面是黑洞的视界。

史瓦西几何论证的基本观点与米歇尔和拉普拉斯发现的相同：像临界周长那么小的星体，从远处看来，必然完全是黑的；一定是我们现在所说的黑洞。尽管基本结果相同，但机理完全不同。

米歇尔和拉普拉斯基于牛顿的空间和时间绝对而光速相对的观点相信，对刚好比临界周长小一点儿的星体，光微粒几乎逃逸了。它们会飞到离星体很远的高度，比任何行星轨道还高；但当它们向上爬的时候，将被星体引力减慢，停在某个尚未达到星际空间的地方，然后调头，被拉回星体。尽管在沿轨道运动的行星上的生物可以通过减速的光看到这颗恒星（对他们来说，它不是黑的），但我们生活在遥远地球上的人却根本看不到它，星光不能到达我们。对我们来说，星体是黑的。

相反，史瓦西的时空曲率要求光总是以相同的普适速度传播，永远也不会变慢。（光速是绝对的，但空间和时间是相对的。）但是，如果光是从临界边界发出的，那么，当它向上经过无限小的距离时，

它的波长必然产生无限大的移动。（波长移动之所以无限大，是因为时间流在视界处是无限膨胀的，而波长总是以与时间膨胀相同的量移动。）波长无限大移动的结果是消耗了所有的光能，因此光也不复存在了！这样，不论行星离临界边界多近，它上面的生物也根本看不见从星体发出的光。

我们将在第7章研究，从黑洞临界面内部看，光的行为是怎样的。我们会发现，光毕竟没有消失，它只不过是不能逃出临界面（黑洞的视界），尽管它仍在以标准的、普适的每秒299792公里的速度向外运动。但是，本书才开头，我们还不能很好地理解这种看似矛盾的行为。现在，我们必须首先建立对他事物的认识，这也是从1916到1960年这几十年间物理学家们所做的。

在20年代和进入30年代后，世界上最有名的广义相对论专家是爱因斯坦和英国天文学家爱丁顿（Arthur Eddinglon）。虽然别人也懂相对论，似爱因斯坦和爱丁顿为这个学科定下了理性的基调。在有人愿意把黑洞当真时，爱因斯坦和爱丁顿不愿意。黑洞就是“味不对”，它们太奇怪了；它们违背了爱因斯坦和爱丁顿关于我们的宇宙应该如何表现的直觉。

爱因斯坦在20年代似乎已经研究过这个问题，却将它忽略了。那时候，没人认为黑洞是什么重要的预言，所以也没有多大必要为此清理这些事情。而且因为别的自然之谜更有趣，更令人困惑，所以爱因斯坦把精力放到了别的地方。

爱丁顿在20年代的方法更天真。他像个业余演员，喜欢普及科学；只要没人把黑洞太当真，它们就是在别人面前摇摆的玩物。于是，我们看到他在1926年的《恒星的内部结构》一书中写道，没有什么可观察的星体能比临界周长的星更致密：“第一，引力的作用会大得连光也不能从它逃逸，光线会像石头落回地球那样落回恒星。第二，谱线的红移会大得连它的谱都不存在了。第三，质量会产生很大

的时空度规的曲率，使空间封闭起来，将星体包在里面，而将我们留在外头（不知那是什么地方）。”第一个结论是光不能逃逸的牛顿式说法；第二个结论是半精确的相对论表达；而第三个结论就是典型的爱丁顿式的夸张了。我们从图3.4的嵌入图已经清楚地看到，当星体像临界周长那么小时，空间弯曲是很强烈的，但还不是无穷大，空间当然也不会卷起来包围星体。爱丁顿大概是知道这一点的，但他的描述很动听，在天真和幽默中把握了史瓦西时空曲率的精神。

我们将在第4章看到，在30年代，要求认真考虑黑洞的压力开始增加了。随着压力的增加，爱丁顿、爱因斯坦和其他一些“一言九鼎”的人物开始明确表示他们对这些怪物的反对。

1939年，爱因斯坦发表了一篇广义相对论的计算文章，作为一个例子，他解释了为什么黑洞不能存在。[7]他的计算分析了人们也许想过可以用来制造黑洞的一类理想物体。那是一个靠引力相互吸引从而聚在一起的粒子集，很像太阳通过引力作用于它的行星而将太阳系团结在一起。爱因斯坦集合的粒子都沿圆轨道围绕一个共同的中心运动，它们的轨道形成一个球面，球面上一端的粒子靠引力吸引着另一端的粒子（图3.5左）。

爱因斯坦想象，让集合越来越小，将实际周长向临界周长压缩。正如我们所预料的，他的计算表明，集合越紧密，球面的引力越强，在球面上运动的粒子为不至被引力拉进去，一定会运动得更快。爱因斯坦证明如果集合小于1.5倍临界周长，引力将非常强大，为避免被它吸进去，粒子不得不比光还跑得快。因为没有东西能比光运动得更快，所以粒子集不可能比1.5倍临界值还小。“至于为什么‘史瓦西奇点’不存在于物理学实体中”，爱因斯坦写道，“这个考察的基本结果说得很清楚了。”

作为对自己观点的支持，爱因斯坦还借助了一个其物质密度为常数的理想星体的内部结构（图3.5右）。这样的星体靠它内部气体的

压力而避免坍缩。史瓦西曾用广义相对论导出了这种星体的完备的数学描述，他的公式表明，如果让星体越来越紧密，那么，为了反抗内部引力强度的增大，星体内部的压力也一定会越来越高。史瓦西公式表明，在星体周长压缩到9/8＝1.125倍临界周长时，中心压力成为无限大。因为真实气体不可能产生真正的无限大的压力（任何其他物质也不能），所以爱因斯坦相信，这样的星体不可能像1.125倍临界周长那么小。[8]

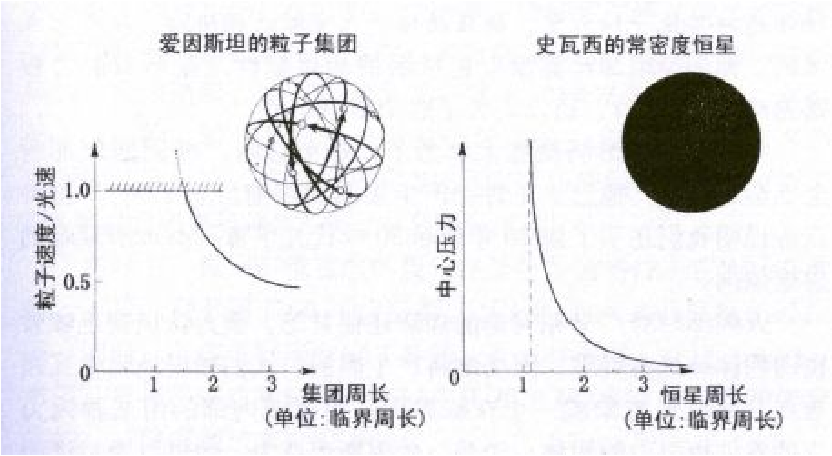



图3.5　爱因斯坦关于不可能存在像临界周长那么小的物体的证明。左：如果爱因斯坦的球状粒子集的周长小于1.5倍临界周长，那么粒子速度一定会超过光速，而这是不可能的。右：如果常密度的星体小于9/8＝1.125倍临界周长，那么星体中心的压强一定会无穷大，这也是不可能的。

爱因斯坦的计算是正确的，但他对结果的解释却错了。他总结的结论，即没有物体能变得像临界周长那么小，更多地是由他反对史瓦西奇点（黑洞）的直觉决定的，而不是根据计算本身。回头来看，我们现在知道，正确的解释是：

爱因斯坦的粒子集和常密度星体不可能达到形成黑洞的致密程度，因为爱因斯坦需要内部的某种力来平衡引力的挤压：在恒星情形，他需要气体压力；在粒子集的情形，他需要粒子运动的离心力。

事实上，在物体很接近它的临界周长时，确实没有什么力量能够抵抗引力的挤压，但这并不意味物体不能这么小。相反，它倒意味着，假如物体真那么小，那么，引力必然超越物体内部其他一切力量，挤压物体产生灾难性的坍缩，从而形成黑洞。爱因斯坦的计算没有包括坍缩的可能性（他所有的方程都忽略了这一点），所以失去了这个结果。

我们今天对黑洞概念太熟悉了，难免会问，“爱因斯坦那时怎么会那么笨？他怎么正好把产生黑洞的坍缩给忘了呢？”这种反应说明我们还不了解20年代和30年代几乎每一个人所具有的思想倾向。

人们那时对广义相对论的理解还很贫乏，没人认识到足够致密的物体必然会坍缩，而坍缩将产生黑洞。人们错误地将史瓦西奇点（黑洞）想象成一个在临界周长附近或内部的由某种内力支持着抗拒引力的物体；于是，爱因斯坦认为，他可以通过证明没有什么由内力支持的东西可以像临界周长那么小，来揭露黑洞的荒谬。

假如爱因斯坦猜想“史瓦西奇点”确实可能存在，那他当然也会认识到坍缩是它们形成的关键，而内力是无关紧要的。但是，他太相信它们不能存在了（它们“味道不对”，非常不对），跟那时差不多所有的同行一样，他的头脑里有一块打不破的石头，遮住了真理。

在怀特（Terence Hanbury White）的史诗小说《过去和未来的国王》里，有一个蚂蚁社会，它们有句格言，“凡不被禁止的事情都是必然的。”物理学定律和真实的宇宙并不是这样的。许多物理学定律允许的事情也是很不可能的，实际上从来没有发生过。一个简单的广为流传的例子是打碎在地板上的鸡蛋自发地重新聚集起来：鸡蛋落到地板上，碎成蛋黄和蛋清，将这个过程拍成电影。然后，将影片倒放，会看到鸡蛋又自发地重新聚在一起，飞离地板。物理学定律允许这样的重聚过程在时间长河里发生，但它从来没有实际发生过，因为太不可能了。

在20年代和30年代，甚至到了40年代和50年代，物理学家的黑洞研究都只在针对一个问题：物理学定律是否允许这种物体存在——而答案是不定的：乍看起来，黑洞似乎是可能的；然后，爱因斯坦、爱丁顿和其他一些人（错误地）论证，它们是不允许的。50年代，当这些论证最终被否定了以后，许多物理学家又来讨论，黑洞也许是物理学定律允许的，但太不可能了（像重组的鸡蛋那样），永远不会实际出现。

实际上，黑洞不像重组的蛋，在某些普通条件下它们是必然的。不过，只是到了60年代后期，当它们的必然的证据已经压倒一切时，大多数物理学家才开始认真面对黑洞。在接下来的三章里，我将叙述这些证据是如何在从30年代到60年代间积累起来的，又曾遇到过多少阻力。

黑洞在20世纪遇到的广泛而且几乎普遍的抵制，与它在米歇尔和拉普拉斯的18世纪受到的热烈欢迎，形成了鲜明的对比。阿尔伯达大学的当代物理学家伊斯雷尔（Werner Israel）深入研究了这段历史，猜测了这种差别的原因。

“我相信，[18世纪对黑洞的接受]不仅是18世纪90年代的革命热情的一个征兆，”他写道，“解释必然是，拉普拉斯的暗星[黑洞]没有给我们对物质的永恒性和稳定性的根深蒂固的信仰带来威胁。相反，20世纪的黑洞对那个信仰却是一个巨大的威胁。”[9]

米歇尔和拉普拉斯都想象，它们的暗星是由与水或土或岩石或太阳密度大致相同（大约每立方英寸[16.39立方厘米]1克）

的物质构成的。以这样的密度，星体若要成为暗星（包围在临界周长内），它的质量就必须比太阳大4亿倍，周长比地球轨道大3倍。这样受牛顿物理学定律支配的星体也许是奇异的，但确实没有威胁到我们关于自然的任何根深蒂固的信仰。如果谁想看看这颗恒星，他只

需要站在邻近的一个行星上看它的光微粒：它们在各自的轨道内上升，然后很快落回恒星的表面。如果谁想要一块恒星组成物质的样本，他只需要飞到恒星的表面，取一块带回地球实验室来做研究。我不知道米歇尔、拉普拉斯和他们那个时代的其他人是不是想过这些事情，显然，如果他们想过，也没有什么理由替自然规律担心，替物质的永恒性和稳定性担心。

20世纪黑洞的临界边界（视界）提出了一个不同的挑战。视界上没有什么地方可以让我们看到任何发出的光。任何落进视界的东西都不可能再逃逸出来；它从我们宇宙消失了，它的消失又给物理学家关于质量和能量守恒的观念带来了严峻的挑战。

“黑洞的历史和大陆漂移（地球上大陆之间的相对漂移运动）的历史有着惊人的相似，”伊斯雷尔写道：“两者的证据到1916年就已经不容忽视了；但两种思想都被一种近乎无理的阻力在半路上阻挡了半个世纪。我相信，两种情况下的基本的心理学原因都是一样的。另外还有一个巧合：两个阻力都大约是在1960年开始崩溃的。当然，两个领域（天体物理学和地球物理学）都获益于战后的技术进步。但更有趣的是，正是在这个年代，苏联的氢弹和人造卫星扫荡了西方科学牢不可破和不容挑战的观念，而且也许还带来了更多的怀疑：在天地之间可能存在着比西方科学所能梦想的更多的东西。”10

[1] 见第10章。

[2] 见图1.3和第2章里关于蒙里迪那和塞罗那岛的故事。

# 第4章　白矮星之谜

爱丁顿和钱德拉塞卡

为大质量星体的死亡而争论；

它们死亡时一定会收缩而产生黑洞吗？

量子力学能拯救它们吗？

那是在1928年，在印度东南部濒临孟加拉湾的马德拉斯市，17岁的印度少年钱德拉塞卡（Subrahmanyan　Chandrasekhar）正在马德拉斯大学沉溺在物理学、化学和数学中。钱德拉塞卡高大英俊，举止大方，踌躇满志。最近，他刚读了索末菲的经典教科书《原子结构和光谱》。现在，他高兴极了，索末菲这位世界大物理学家从他在慕尼黑的家来马德拉斯访问了。

钱德拉塞卡急切地想接近索末菲，他来到他所在宾馆的房间，请他接见。索末菲答应了，约他几天后来。

约好的那天，以为掌握了现代物理学的钱德拉塞卡满怀骄傲和自信，来到索末菲的房间，敲了门，索末菲礼貌地请他进来，询问了他的学习情况，然后，向他泼了点儿凉水。他解释说：“你学的物理是过去的东西，在我的书写好以后的五年里，物理学有了很大的变化。”他接着介绍了物理学家在对统治微观领域的规律的认识中发生的革命，在这个分子、原子、电子和质子的领域里，牛顿定律以相对论没有料到的方式失败了，取代它们的是一组根本不同的物理学定律——量子力学的定律，[1]因为它们处理的是物质粒子（“量子”）的行为

（“力学”）。尽管新的量子力学定律才两岁，但在解释原子和分子如何运动方面，已经获得了巨大的成功。

钱德拉塞卡在索末菲的书中读到了新定律的原始形式。但索末菲告诉他，原始量子定律是不令人满意的。虽然对像氢那样的简单原子和分子来说，它们与实验符合得很好，但不能解释更复杂的原子和分子的行为，而且也不能在逻辑上同其他物理学定律一致地吻合。它们不过是一些难看的、特设的计算法则的大杂烩。

定律的新表达尽管在形式上太离奇了，看起来却更有希望。它解释了复杂的原子和复杂的分子，而且似乎同其他物理学也吻合得非常好。

钱德拉塞卡仔细地听着，出神了。

## 量子力学与白矮星的内部结构

告别时，索末菲给了钱德拉塞卡一篇他刚写好的论文校样，文章推导了决定被挤压在一个小体积（如金属）内的大电子集团行为的量子力学定律。

钱德拉塞卡如痴如醉地读索末菲的校样，读懂了，然后又用了好些天在图书馆学习他能找到的所有与此有关的论文。特别令他感兴趣的是英国物理学家福勒（R.H.Fowler）的一篇题为“论致密物质”的文章，发表在1926年12月10日出版的《皇家天文学会月报》上。[1]福勒的文章将钱德拉塞卡引向了一本更令人着迷的书，著名英国天体物理学家爱丁顿的《恒星的内部结构》，[2]在这本书里，钱德拉塞卡发现了如何描述白矮星的秘密。

白矮星是天文学家通过望远镜发现的一种星体，白矮星的神秘在于它内部物质的极高密度远大于人类曾经遇到过的密度。钱德拉塞卡

在打开爱丁顿的书时，还无从知道它，但为了揭示这个高密度的秘密，他和爱丁顿最后不得不面对致密星体存在的可能，这些星体死后，会收缩形成黑洞。

“白矮星可能很多，”钱德拉塞卡从爱丁顿的书中看到，“我们只有限地知道3个，但它们都在离太阳很近的距离内……其中最著名的是[正常星]天狼星的伴星”，叫天狼B。天狼星和天狼B是距地球第6和第7近的两个星体，8.6光年远，而且，天狼星还是我们天空中最亮的恒星。天狼B像地球绕太阳那样围绕天狼星旋转，但它转一周需要50年，而地球只需要1年。

爱丁顿讲了天文学家如何根据望远镜的观测来估计天狼B的质量和周长。它的质量是0.85个太阳质量，周长是118 000千米。这意味着天狼B的平均密度为每立方厘米61 000克——水密度的61 000倍，大约正好每立方英寸1吨。“这个方法我们已经知道好多年了，我想很多人都想过，加上‘这个荒谬的’结论也是恰当的。”大多数天文学家不可能认真地看待一个比在地球上所遇到的大那么多的密度——假如他们知道更多的现代天文观测所揭示的那些事实（1.05个太阳质量，31 000千米的周长，每立方厘米400万克或每立方英寸60吨的密度），他们会认为更荒谬。见图4.1。

爱丁顿接着描述了巩固这个“荒谬”结论的一个关键的新观测。如果天狼B的密度确实是水的61 000倍，那么根据爱因斯坦的引力定律，从它的强引力场中“爬出来”的光将向红端移动十万分之六——比从太阳发出的光的红移大30倍，从而更容易测量。红移的预言似乎在1925年爱丁顿的书即将出版前就经过了威尔逊山天文台（坐落在加利福尼亚帕萨迪纳的一个山顶上）的亚当斯（W.S.Adams）的检验和证实。[2]爱丁顿写道，“亚当斯教授一石二鸟，实现了爱因斯坦广义相对论的新检验，还证实了我们的猜想：比天狼B致密2 000倍的物质不仅是可能的，而且在宇宙中确实出现了。”

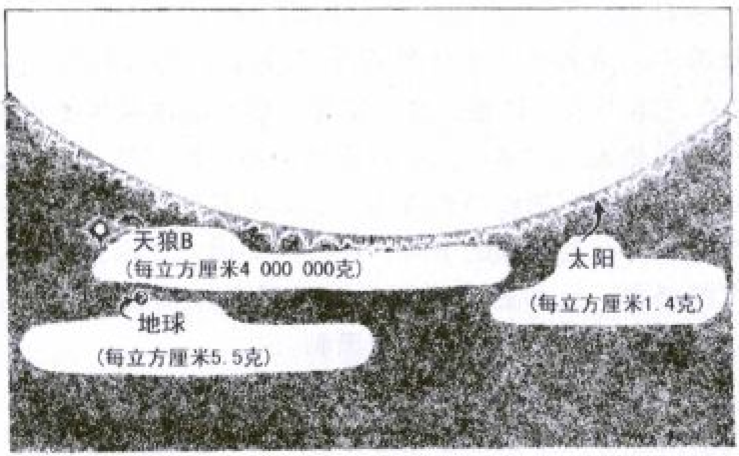



图4.1　太阳、地球和白矮星天狼B的大小和平均密度比较（现代数值）

在爱丁顿的书里，钱德拉塞卡还看到了星体（如太阳和天狼B）的内部结构是如何靠内部压力与引力的平衡来维持的。压缩与膨胀的平衡，可以通过一个类比来理解（虽然这不是爱丁顿的方法）：挤压你手中的气球（图4.2左），你双手向内挤压的力正好被气球空气向外的压力所平衡——空气压力是由气球内部的空气分子撞击气球的橡皮壁而产生的。

对星体（图4.2右）而言，与你挤压的双手类似的，是星体外层物质的重量，而与气球内空气类似的是星体在那一层以内的球形物质。外层与内球的边界可以选在你喜欢的任何地方——星体内部1米深、1千米深、1000千米深，都可以。不论边界选在什么地方，它都必须满足这样的要求：挤压内球的外层星体物质的重量正好被内球分子撞击外层的压力所平衡。这种在星体内处处存在的平衡，决定了星体的结构，也就是说，决定了星体的压力、引力和密度是如何从星体表面向中心变化的。

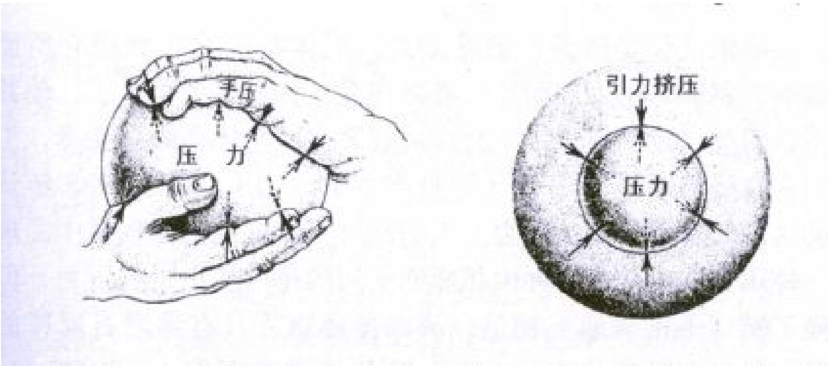


图4.2　左：双手的挤压与气球内部压力之间的平衡。右：星体外层物质的引力挤压（重量）与层内球体物质膨胀之间的平衡。

爱丁顿的书还讲述了那时所知道的关于白矮星结构的一个恼人的困惑。爱丁顿相信——实际上在1925年所有的天文学家都相信——白矮星物质的压力一定像气球那样是由热产生的。热使物质的原子以很高的速度在星体内部四处飞动，相互碰撞，并撞击星体外层与内球间的界面。如果用“宏观的”观点看（对探测单个原子来说它太粗了），那么，我们能够测量的只是一个总的撞击力，例如所有原子撞击每平方厘米界面的力，这个总力就是星体的膨胀压力。

星体通过向空间发出辐射而冷却，它的原子将随之而越飞越慢，原子的压力也下降，从而外层星体的重量会将内球挤压到更小的体积。然而，球体的压缩又将星体加热，增大它的膨胀，于是又能达到一个新的压缩-膨胀平衡——这时星体比原先小一点。这样，随着向星际空间辐射热量而冷却，星体必然会逐渐地收缩变小。

这种逐渐收缩的过程如何结束呢？天狼B的最终命运是什么？最显然（但是错误）的答案是，星体将一直收缩到它所能形成的黑洞那么小，然而这个答案在爱丁顿看来太讨厌了，他甚至不曾想过。他断言，惟一合理的答案是，星体最终会变冷，但支持它自身的不是热压力（即热产生的压力），而是1925年认识的惟一的另一类型的压力：人们在岩石那样的固体物质中发现的一种压力，也就是一种由相邻原

子间的排斥所产生的压力。但是爱丁顿（不正确地）相信，假如星体物质具有像岩石那样的密度，每立方厘米几克——比天狼B表现的密度小10 000倍，那么这样的“岩石压力”就是惟一可能的。

沿着这样的思路，爱丁顿遇到了一个疑惑。星体为了重新扩张到岩石的密度从而能在冷却后支撑自身，它不得不做很大的功来克服自身的引力，而物理学家不知道星体内有什么能量供应能满足这样的功。“想想看，一个物体在不断地失去热量却没有足够的能量变冷！”爱丁顿写道，“这是一个很离奇的问题，至于实际会发生什么事情，我们可以想出很多建议。这个困难还不一定是致命的，我们在这儿就将它抛在一边。”

钱德拉塞卡在福勒1926年“论致密物质”的文章里发现了这个1925年疑难问题的解决办法，解决的基础在于他认识到爱丁顿所用的物理学定律失败了。那些定律必须用新的量子力学来代替，量子力学不是将天狼B和其他白矮星的内部压力归因于热，而是将它们归因于一种新的量子力学现象：电子的退化运动，或者叫电子简并[3]。

电子简并有点儿像人的幽闭。当物质被挤压到比岩石高10 000倍的密度时，围绕各原子核的电子云被压缩10 000倍，从而每个电子被限制在比它原来可以活动的空间体积小10 000倍的“格子”内。因为只有这么小的活动空间，所以电子就像幽闭的人情不自禁地颤栗，开始在小格子里高速地飞来飞去，以极大的力量撞击格子里的相邻电子。这种物理学家所说的退化的（简并）运动不可能靠物质冷却来阻止，没有什么东西可以令它停止，它是被量子力学定律强迫作用在电子上的，即使物质处在绝对零度，它仍然存在着。

简并运动是牛顿物理学家做梦也没想到的一种物质特性的结果，这种特性叫作波粒二象性：根据量子力学，每种粒子的行为有时像波，而每种波的行为有时像粒子。这样，波和粒子其实是一个东西，一个有时像波有时像粒子的“东西”，见卡片4.1。

用波粒二象性的概念，电子简并是很容易理解的。当物质被压缩到高密度时，物质内部的每个电子都限制在被相邻格子挤压成的极小格子里，部分表现出波的样子。电子的波长（波峰间的距离）不能大于电子的格子，否则波就会超过格子。这时候，波长很短的粒子必然是高能的。（一个普通的例子是与一个电磁波相联系的粒子，即光子。X射线光子的波长远小于可见光光子的波长，结果，X射线光子的能量远比可见光光子大。高能的X射线光子能够透过人体的骨肉。）

在非常致密的物质中，电子的短波长和相应的高能量意味着快速运动，就是说电子一定在格子里四处飞动，不规则地快速地变来变去，一半是粒子，一半是波。物理学家说电子是“简并的”，他们将这种由电子的高速不规则运动产生的压力称为“电子简并压力”。没有办法摆脱这个简并压力，它是电子限在小格子里不可避免的结果。而且，物质密度越大，格子越小，电子波长越短，能量越高，运动越快，从而简并压力越大。在寻常密度的寻常物质中，电子简并的压力太小了，没有人发现过，但在白矮星的巨大密度下，它是很显著的。

卡片4.1　波粒二象性简史

还在伊萨克·牛顿的时代（17世纪末），物理学家就争论过，光是粒子还是波组成的？牛顿在这个问题上虽然犹豫不决，但还是倾向粒子，称它们为微粒，而惠更斯（Christiaan Huygens）主张是波。19世纪初以前，牛顿的粒子观点一直占统治地位。后来发现，光可以发生干涉（第10章），物理学家就转向了惠更斯的波动观点。19世纪中叶，麦克斯韦以他统一的电磁定律为波动描述奠定了坚实的基础，物理学家那时就认为，问题最终解决了。不过，这是在量子力学以前的事情。

19世纪90年代，普朗克在热物体发出的光的谱线形态中，发现物理学家在光的认识上可能忽略了什么东西。1905年，爱因斯坦找到了那失去的东西：光的行为有时像波，有时像粒子（现在叫光子）。爱

因斯坦解释，当它与自身干涉时，它像波；但在光电效应中，它像粒子：当一束微光照在金属片上时，光每次从金属片打出一个电子，就像每一个光的粒子（单个光子）打在电子上，然后一个一个地将它们从金属表面打出来。爱因斯坦根据电子的能量推测，光子的能量总是与光的波长成反比。这样，光子和光的波动性质就交织在一起了，波长不可抗拒地与光子的能量联系在一起了。爱因斯坦关于光的波粒二象性的发现和他开始在这个发现的基础上建立的原始的量子力学的物理学定律，在1922年为他赢得了1921年度的诺贝尔奖。

尽管爱因斯坦几乎是一手建立了广义相对论，但在量子力学定律——关于“小东西的天地”的定律，他只是众多贡献者中的一个。爱因斯坦发现光的波粒二象性时，还没认识到电子和质子也能有时表现像粒子，有时像波。直到20年代中期，还没有人认识到这一点，后来，德布洛意（Louis de Broglie）将它作为一个猜想提出来，然后，薛定锷（Erwin SchrΦdinger）将它作为一个完整的量子力学定律的基础，在这个定律中，电子是一种几率波。那是关于什么的几率呢？粒子位置的几率。这些“新”的量子力学定律在我们这本书里没有多大关系（在解释电子、质子、原子和原子核如何表现上，它已经取得了巨大的成功）。不过，它们的一些性质会越来越重要。在本章里，重要的性质是电子简并。

爱丁顿写书的时候，还没有谁预言过电子简并，也就没有可能正确地计算在压缩到天狼B那样的超高密度时岩石或其他材料会产生什么反应。现在有了电子简并的定律，这样的计算成为可能，而且福勒在1926年的文章里已经想到并实现了。

根据福勒的计算，因为在天狼B和其他白矮星内的电子被压缩到那么小的格子里，所以它们的简并压力远大于它们的（热运动引起的）热压力。相应地，当天狼B冷却下来后，它微弱的热压力将消失，但它们巨大的简并压力将保留并持续地支撑着它抵抗引力。

这样，爱丁顿的白矮星疑难的解决在于两个方面：（1）天狼B对抗它自身的引力，所依靠的并不是人们在新量子力学出现之前所想的热压力，而是简并压力。（2）天狼B冷却后，不必再膨胀到岩石的密度来维持自己；相反，它在每立方厘米400万克的密度下的简并压力足以让自己继续维持下去。

在马德拉斯图书馆里读着这些东西，学着它们的数学公式，钱德拉塞卡入迷了。这是他头一回接触现代天文学，他在这儿看到20世纪两大物理学革命的深刻结果会聚到一起了：爱因斯坦的广义相对论，因它关于空间和时间的新观点，在来自天狼B的光线的引力红移上表现出来了；而新的量子力学，因它的波粒二象性而产生了天狼B的内部压力。天文学是一块肥沃的土地，年轻人能在那儿尽情地耕耘。

钱德拉塞卡在马德拉斯继续求学，进一步探讨天文学宇宙的量子力学结果。他甚至将自己的思想写成一篇小文章，寄给英国他从没见过的福勒，福勒想办法让它发表了。

终于，1930年19岁那年，钱德拉塞卡在印度完成了相当于美国学士学位的学业，在7月的最后一个星期，他登上了驶向英国的轮船。他已经被录取为剑桥大学的研究生，那里是他心中的英雄福勒和爱丁顿的家乡。

[1] 关于量子力学定律的清晰讨论，见Heinz Pagels的The Cosmic Code（Simon and Schuster，1982）.（中译本是《宇宙密码》，朱栋培、陈宏芳译，中国科学技术大学出版社，1988.——译者）

[2] 得到一个结果，认为它正是我们想得到的结果，这在精密测量里是容易而又危险的，亚当斯的红移测量就是一个例子。他的结果符合预言，但由于天文学家对天狼B的质量和周长的估计的误差，预言却是严重错误的（小了5倍）。[3]。

[3] “degenerate”（退化）一词在这儿并不是它在“道德败坏”（“moral degeneracy”）里的原义（最低的道德水平），而是说电子达到它们的最低可能的能级。

## 极大质量

从马德拉斯到南安普顿的海上18天里，钱德拉塞卡摆脱了常规的学习和考试。许多个月以来，他第一次有机会静静地思考物理学。大海的单调有助于思想，而钱德拉塞卡的思想是很丰富的，真的很丰富，还将为他赢得诺贝尔奖，不过那是54年以后，他成功地使他的思想得到世界天文学界认同之后的事情了。

上船之后，钱德拉塞卡的头脑里浮现着白矮星、爱丁顿的疑惑和福勒的解决办法。福勒的方法几乎肯定是对的，没有别的办法。不过，福勒没有完全认识到白矮星中简并压力与引力之间平衡的细节，也没有计算星体最终的内部结构——当我们通过它的表面到达中心时，它的密度、压力和引力是如何变化的。他遇到一个有趣的挑战，可以帮他消除长途旅行的无聊。

作为发现星体的结构的工具，钱德拉塞卡需要知道下面这个问题的答案：假定白矮星物质已经被压缩到某一密度（例如，每立方厘米100万克），将它再压缩（就是说，减小它的体积，增大它的密度）1%，则它将通过提高压力来反抗这点增大的压缩，那么它的压力会增大多少个百分点呢？物理学家用绝热指数来称1%的压缩所产生的压力增大的百分比。在本书中，我将用一个更形象的名字，压缩阻抗或者简单说，阻抗。（这个“压缩阻抗”不应与“电阻”相混，它们是完全不同的概念。）

钱德拉塞卡解出压缩阻抗，是通过一步步地检验白矮星物质密度每增加1%所产生的结果：电子格子的减小，电子波长的减小，电子能量和速度的增加以及最终电子压力的增加。[4]结果很清楚：每1%的密度增加产生5/3个百分点（1.667%）的压力增加，从而白矮星物质的阻抗为5/3。

在钱德拉塞卡这次旅行的许多年前，天文学家已经计算了物质压缩阻抗与星体深度无关的星体内部引力与压力相平衡的细节——就是说，压力与密度彼此同步增加的星体，在越来越深入它的内部时，每个百分点的密度增加总是伴随着同一固定百分比的压力增加。结果产生的星体结构情形都包含在爱丁顿的《恒星的内部结构》一书里，钱德拉塞卡把书带上了船，因为他太珍爱它了。这样，当钱德拉塞卡发现白矮星物质有5/3的与密度无关的压缩阻抗时，他满意了。现在他可以直接深入爱丁顿的书去发现星体的内部结构：星体的密度和压力从表面到中心的变化方式。

钱德拉塞卡将爱丁顿书中的公式与他自己的公式联系起来，发现天狼B中心的密度为每立方厘米360　000克（每立方英寸6吨），电子简并运动速度为光速的57%。

这样的电子速度大得惊人。钱德拉塞卡像他之前的福勒一样，用量子力学定律计算了白矮星物质的阻抗，但忽略了相对论效应。然而，当任何物体以近乎光的速度运动时，即使粒子服从量子力学定律，狭义相对论的效应也必然变得重要了。在57%的光速，相对论效应可能还不太大，但引力更强的更致密的白矮星需要更大的中心压力来维持自己，它的电子的随机速度也将相应地更大。对这样的白矮星，相对论效应当然不能忽略。所以，钱德拉塞卡回到他分析的出发点，计算白矮星物质的压缩阻抗，这回他决心把相对论效应包括进来。

为将相对论纳入计算，需要将狭义相对论的定律与量子力学定律融合起来——这个融合，理论物理学的伟大头脑们那时刚开始考虑。一个人在船上，又刚从大学毕业，钱德拉塞卡不可能实现完全的融合，但是，他能达到的融合足以揭示出高速电子的主要效应了。

量子力学坚持，当已经致密的物质再压缩一点，使每个电子的格子比原来更小时，电子的波长必然减小，相应地，简并运动的能量必

然增大。然而，钱德拉塞卡认识到，增加的电子能量在性质上是不同的，它依赖于电子的运动是低于光速还是接近光速。假如电子运动慢，那么像平常一样，能量的增加意味着更快的运动，也就是说，电子将具有更高的速度。然而，假如电子接近光速运动，那么它的速度就没有办法增得更高，（否则，它就将超过光速了！）所以能量的增加有不同的形式，是我们在日常生活中所不熟悉的：增加的能量变成惯性；就是说，它增加了电子加速的阻力——电子表现为似乎更重了一点。增加的能量的这两种不同命运（增加速度或者增加惯性）产生不同的电子压力的增加，也就是不同的压缩阻抗——钱德拉塞卡推出：在低电子速度时，压缩阻抗为5/3，与他以前的计算相同；在高速时，压缩阻抗为4/3。

将相对论性简并物质（也就是简并电子以近乎光的速度运动的致密物质）的4/3的阻抗与爱丁顿书中给的公式联合起来后，钱德拉塞卡导出了高密度、大质量白矮星的性质。答案令人震惊；高密度物质要让自己对抗它的引力是困难的——星体只有在小于1.4个太阳质量时，才可能与挤压它的引力相抗衡。这意味着没有哪颗白矮星的质量可以超过1.4个太阳！

钱德拉塞卡的天体物理学知识有限，这个奇怪结果的意义令他感到疑惑。他一次次检查了计算，但找不出错误。于是，在旅行的最后几天，他写了两篇稿子想发表。在一篇文章里，他描述了关于天狼B那样的低质量低密度白矮星结构的结论。在另一篇文章里，他简要地解释了他的结果：没有一颗白矮星能比1.4个太阳更重。

钱德拉塞卡来到剑桥时，福勒正在国外。9月，福勒回来了，钱德拉塞卡急切地去他的办公室，把两篇稿子交给他。福勒赞同第一篇，将它寄给《哲学杂志》发表。但第二篇，关于白矮星最大质量的那一篇，却令他迷惑。他不理解钱德拉塞卡关于没有白矮星能比1.4个太阳更重的证明。他那时是物理学家，不是天文学家，于是他去请他的同事，著名天文学家米尔恩（E.A.Milne）帮着看看，当米尔恩也

不明白这些证明时，福勒就没有拿它去发表。

钱德拉塞卡很着急。他到英格兰3个月了，文章在福勒那儿也搁了2个月，他没有那么多时间等待发表。最后，钱德拉塞卡被惹火了，他放弃了文章在英国发表的努力，把稿子寄给了美国的《天体物理学杂志》。

几星期后，芝加哥大学的编辑回信了：手稿已经寄给美国物理学家埃卡特（Carl Eckart）审查。在稿子中，钱德拉塞卡不加解释地叙述了他的相对论和量子力学计算的结果：在极高密度下的压缩阻抗为4/3。4/3的阻抗对限制白矮星质量是有基本意义的。假如阻抗大于4/3，那么白矮星可以要多重有多重——埃卡特认为它应该更大一些。钱德拉塞卡回了信，回信有4/3阻抗的数学推导；埃卡特细读过后，相信钱德拉塞卡是对的，同意发表他的论文。[5]在钱德拉塞卡写成整整一年后，论文终于发表了。[1]

天文学界对论文没什么反应，似乎没人感兴趣。所以，为了完成博士学位，钱德拉塞卡只好转到别的更容易被接受的研究上去了。

3年后，钱德拉塞卡成了博士，来到俄罗斯同苏联科学家交流研究思想。在列宁格勒，年轻的亚美尼亚天体物理学家安巴楚勉（Viktor Amazapovich Ambartsumian）告诉钱德拉塞卡，世界的天文学家不会相信他对白矮星质量的限制，除非他能根据物理学定律计算有代表性的白矮星样本的质量，并具体说明它们都在他宣布的极限以下。他还指出，钱德拉塞卡分析了相当低的密度下5/3阻抗的白矮星和极端高密度下4/3阻抗的白矮星，这是不够的。他还需要分析大量的密度在二者之间的白矮星样本，并证明它们的质量也总是低于1.4个太阳。一回到剑桥，钱德拉塞卡就担起了安巴楚勉的挑战。

钱德拉塞卡需要的一个基础是白矮星物质在从低到极高的整个密度范围内的状态方程。（物理学家说的物质的“状态”，指的是物质的

密度和压力——或等价地说，物质的密度和压缩阻抗，因为我们可以根据阻抗和密度来计算压力。“状态方程”指的是阻抗与密度间的关系，也就是阻抗作为密度的函教。）

1934年下半年，在钱德拉塞卡接受安巴楚勉的挑战时，白矮星物质的状态方程已经知道了，这要归功于英国利兹大学的斯托纳和爱沙尼亚塔图大学的安德森（Wilhelm Anderson）的计算。[7]斯托纳-安德森状态方程表明，当白矮星物质的密度被挤压得越来越高，从非相对论的低密度和低电子速度区域进入相对论的高密度和近光速的高电子速度区域，物质的压缩阻抗光滑地从5/3降到4/3（图4.3左）。阻抗的变化不会比这更简单了。

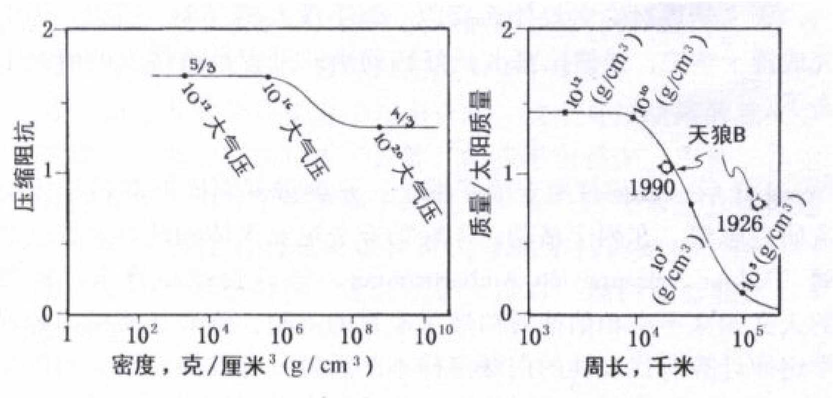



图4.3　左：白矮星物质的斯托纳-安德森状态方程，也就是物质的密度与它的压力之间的关系。水平方向是物质被压缩到的密度，垂直方向是物质的阻抗（1%的密度增加所伴随的压力增加的百分点）。曲线上的点是以地球大气压为单位的挤压力（等于内部压力）。右：钱德拉塞卡用爱丁顿的布伦瑞克机械计算器计算的白矮星的周长（水平方向）和质量（垂直方向）。曲线上的点是以克每立方厘米为单位的星体中心的物质密度。[8]

为了回安巴楚勉的挑战，钱德拉塞卡不得不将状态方程（阻抗对密度的依赖性）与引力和压力力相平衡的星体定律联合起来，从而得到一个描述星体内部结构——也就是描述密度随到星体中心距离而变化的微分方程。[2]然后，他还得为中心密度范围从低到极高的十多个星体求解这个微分方程。只有对每颗星都解出了微分方程，他才能知

道星体的质量，才知道它是否小于1.4个太阳质量。

对中心密度低和极高的星体，钱德拉塞卡在船上就研究过，他发现了微分方程的解和爱丁顿书中得到的星体结构。但是，对中间密度的星体，爱丁顿的书没什么帮助，而且_，不论费多大气力，钱德拉塞卡也没能用数学公式导出方程的解。数学太复杂了，没有别的办法，只能在计算机上数值求解他的微分方程。

1934年那个时候的计算机跟90年代的大不相同。它们更像最简单的袖珍计算器：一次只能进行两个数的乘法，使用者先得用手输入数据，然后摇动曲柄。曲柄带动复杂的齿轮系统进行乘法运算，给出答案。

这种计算机很贵，很难得到，不过爱丁顿行一台“布伦瑞克”（Braunschweigen），大小像90年代初期的个人台式电脑。钱德拉塞卡这时同这位大人物已经很熟了，所以他去爱丁顿那儿向他借计算机。爱丁顿那时正为白矮星问题与米尔恩争得不可开交，很想看看白矮星结构的全部情形，所以他让钱德拉塞卡把布伦瑞克搬到他在三一学院的宿舍里去。

计算冗长而令人生厌。每天晚饭后，爱丁顿（那时是三一学院的老师）都会爬上楼梯，来到钱德拉塞卡的小屋里，看计算怎么运行，能为他带来什么令人鼓舞的东西。

好多天以后，钱德拉塞卡终于算完了。他回答了安巴楚勉的挑战。对10个有代表性的白矮星，他计算了每一个的内部结构，然后根据内部结构计算了星体的总质量和周长。所有质量都像他顽强期待的那样小于1.4个太阳。而且，当他把星体质量和周长点在图上并“将点联结起来”时，他得到了一条光滑曲线（图4.3右；也可参见卡片4.2），天狼B和其他已知白矮星的观测质量和周长都很好地符合这条曲线。（随着现代天文观测的进步，符合情况更好了。注意图4.3中

天狼B的质量和周长的1990年最新数值。）结果令人自豪，想到全世界的天文学家将终于接受他说的白矮星不能比1.4个太阳更重，钱德拉塞卡感到无比的幸福。

特别令他满足的还是他有机会向伦敦的皇家天文学会报告这些结果。钱德拉塞卡的报告时间安排在1935年1月11日星期五。照惯例，在会议开始之前，会议议题的细节是保密的。不过，钱德拉塞卡的朋友、学会助理秘书威廉（Kay Williams）小姐总会偷偷地提前将会议内容告诉他。星期四傍晚，他收到了会议内容的邮件，惊讶地发现紧跟在他自己讲话后面的是爱丁顿关于“相对论性简并”的讲话。钱德拉塞卡有点儿生气。在过去几个月里，爱丁顿每星期至少来一次，看他的工作情况，还读过他正在写的论文草稿，却一句也没提他自己在这个课题上的任何研究！

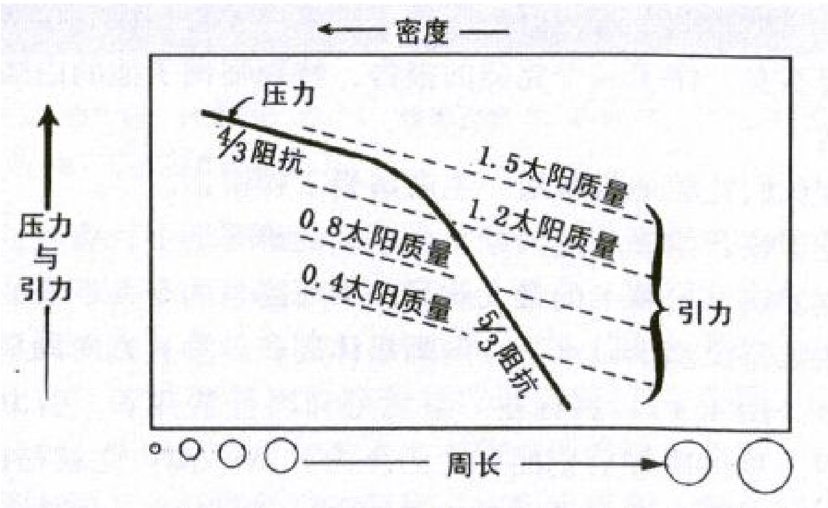
钱德拉塞卡抑制着愤怒下楼去晚餐，爱丁顿也在那儿，在高桌上用餐。钱德拉塞卡于是坐在另一处，一句话也不说，因为礼仪告诉他，像爱丁顿那样的大人物，你认识他，而且他又表示了对你的工作感兴趣，那么你就没有权力因为这样的事情去打扰他。

晚饭后，爱丁顿自己把钱德拉塞卡找出去说，“我已经请斯马特（Smart）明天给你半个小时的时间，而通常只有15分钟。”钱德拉塞卡谢了，还等他说他自己的讲话，但爱丁顿却找借口走了。钱德拉塞卡恼火之外，现在更感到焦急和痛苦。

卡片4.2　白矮星的质量和周长的解释

为了定量地认识为什么白矮星有如图4.3所示的质量和周长，我们来看下面的图。它表示了作为星体周长（画在右边）和密度（画在左边）的函数的白矮星内部的平均压力和引力（画在上边）。如果星体被压缩，那么它的密度会增加，周长会减小（图中向左运动），星体压力将沿实线上升，在压缩阻抗为5/3的低密度上升较陡，在阻抗

为4/3的高密度上升较缓。同样的星体压缩还导致星体表面向中心运动，从而沿虚线增大了星体内部引力的强度。引力的增长率类似于4/3的阻抗：每1个百分点的压缩对应着4/3个百分点的引力强度的增加。图中画出了几条引力虚线，每条线对应一个星体质量，因为星体质量越大，它的引力越强。




在每颗恒星（例如，1.2个太阳质量的星体）的内部，引力和压力必须相互平衡，因此这颗星必然处在标有“1.2太阳质量”的引力虚线与压力实线的交点上，这个交点决定了恒星的周长（标在图的底部）。如果周长更大，则恒星的引力虚线将在压力实线之上，引力超过压力，星体向内坍缩；假如周长较小，压力将超过引力，星体发生爆炸。

几条虚线与实线的交点对应于图4.3右图的平衡白矮星的质量和周长。对小质量的恒星（最低的那条虚线），交点处的周长较大；质量更大的恒星（高处的虚线）周长较小。1.4个太阳质量的星体，就没有任何交点，引力虚线总是在压力实曲线的上面，所以引力总比压力大，不论星体的周长多大，引力都迫使它坍缩。

[1] 同时，斯托纳（Edmund C.Stoner）也独立推导并发表了白矮星最大质量的存在，但他的推导不如钱德拉塞卡的令人信服，因为他假定了星体在整个内部具有不变的密度。[6]

[2] 微分方程将不同的函数和它们的变化率，也就是将函数和它们的“导数”结合在一个单独的公式里。在钱德拉塞卡的微分方程中，函数是星体的密度和压力以及它的引力强度，它们都是到星体中心距离的函数。微分方程是这些函数与它们在从中心穿过星体向外运动时的变化率之间的关系。“求解微分方程”的意思是，“根据这个微分方程计算函数本身”。

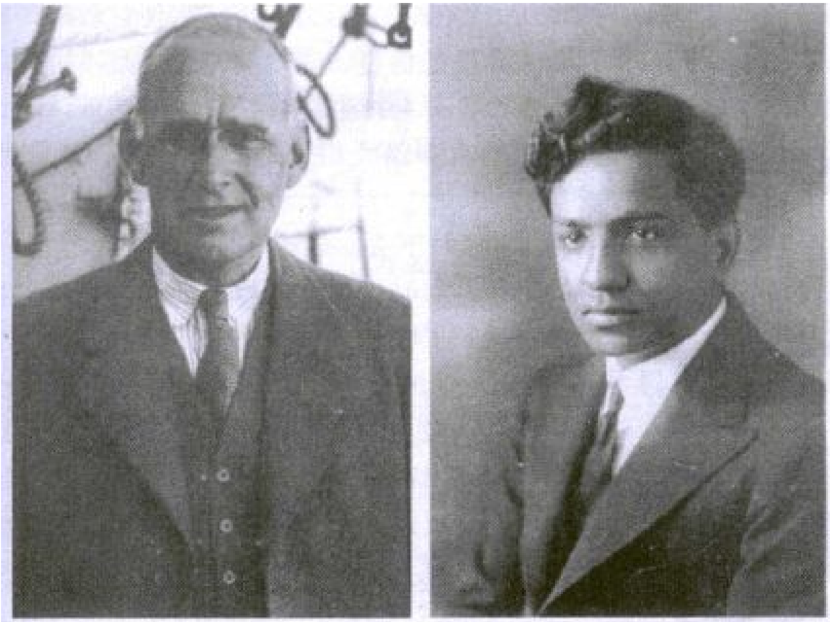
# 论战

第二天早晨，钱德拉塞卡坐火车到伦敦，然后乘出租车来到伯林顿宫的皇家天文学会。在他和朋友麦克里（Bill McCrae）等着会议开幕时，爱丁顿走过来，刚看了议题的麦克里问他，“爱丁顿教授，我们该怎么理解您说的‘相对论性简并’？”爱丁顿转向钱德拉塞卡，回答说，“那会令你吃惊的。”说着就走了，令钱德拉塞卡更加焦虑不安。

会终于开始了。学会主席宣布了许多事情，许多天文学家又做了不同性质的发言，时间拖长了。最后才轮到钱德拉塞卡。他抑制着不安，作了一个完美的报告，特别强调了他的白矮星的最大质量。

会员们礼貌地鼓掌后，主席请爱丁顿讲话。

爱丁顿开始平和地回顾了白矮星的研究历史，然后，他激动地讲述了钱德拉塞卡的最大质量结果所隐含的令人不安的事实：在钱德拉塞卡以水平方向画星体周长、垂直方向画星体质量的图中（图4.4），只有在一组质量和周长条件下，引力能被非热压力（星体冷却后仍然存在的压力）所平衡：这就是白矮星。在钱德拉塞卡白矮星曲线的左边区域（阴影区，小周长星体），星体的非热简并压力完全超过了引力。简并压力将使区域内的任何星体发生爆炸。在白矮星曲线右边的区域（白区，较大周长的星体），引力完全超过了星体的简并压力。任何一颗处在这个区域的冷星都将在引力挤压下立刻发生坍缩。



左：A.S.爱丁顿，1932年。右：S.钱德拉塞卡，1934年。[左：UPI/Bett-man提供；右：钱德拉赛卡提供。

]

太阳能存在于白区只是因为它现在还很热；它的热（热产生的）压力设法平衡了引力。然而，当太阳最终冷却下来时，热压力将消失，从而不可能继续维持自身。引力将迫使它收缩得越来越小，将它的电子挤到越来越小的格子里去，直到它们最后具有足够的简并压力（非热压力）来抵抗引力，阻止压缩。在这个压缩“死亡”的过程中，太阳质量将近似保持为常数，但它的周长会减小，所以它将沿着图4.4中的水平线向左移动，最终停在白矮星曲线上——那是它的归宿。太阳将成为一颗白矮星永远停留在那里，逐渐冷却，成为黑矮星——一颗冷的、暗淡的固态星体，约地球大小，但质量和密度比地球大100万倍。

太阳的这个最终命运似乎令爱丁顿很满意。比钱德拉塞卡的1.4

个太阳质量的白矮星极限质量更大的星体——例如，天狼星，天狼B的伴星，2.5个太阳质量——就没有这样令人满意的归宿。假如钱德拉塞卡是正确的，这样的星体就永远不可能像太阳那样平静地死亡。当它向空中发出的辐射带走了足够的热量而开始冷却时，它的热压力将衰减，引力的挤压将使它收缩得越来越小。像天狼星那么大质量的恒星，非热简并压力是不可能阻挡这样的收缩的。这一点在图4.4中看得很清楚，在那里，阴影区没有延伸到能与天狼星的收缩路线相交的高度。爱丁顿发现，这个预言令人不安。

“恒星将不得不持续地辐射下去，收缩下去。我想，它会一直辐射收缩到几公里的半径，那时引力会变得很强大，足以平息这些辐射，而恒星也最终找到了安宁。”（用90年代的话说，它必然形成黑洞。）爱丁顿告诉他的听众，“钱德拉塞卡博士以前得到过这个结论，在最近的一篇文章里他又特别强调了这一点。在同他讨论的时候，我被迫作出这样的结论：这几乎是相对论性简并公式的一个反证。可能会出现许多事件来挽救恒星，但我想的不仅是这样的保护。我想，应该存在一个自然律来阻止恒星那么荒谬的行为！”[9]

接着，爱丁顿论证钱德拉塞卡结果的数学证明不能令人相信，因为它的基础是在没有充分根据的条件下人为地将狭义相对论与量子力学揉在一起。“这样的产儿，我想不会是合法婚姻的结果。”爱丁顿说，“我自己满意的是[如果揉合是对的]，相对论修正将获得补偿，于是我们回到‘平常的’公式”（也就是回到5/3的阻抗，它允许白矮星有任意大的质量，从而压力能够阻止天狼星在图4.4中假想虚线上的收缩）。然后，爱丁顿大概讲了他认为狭义相对论和量子力学应该怎样融合，融合的方法与钱德拉塞卡、斯托纳和安德森用过的都不同。爱丁顿宣称，这将使所有的恒星都摆脱黑洞的命运。

钱德拉塞卡惊呆了。他从没想到他的工作会遭到这么大的攻击。爱丁顿为什么不事先跟他讨论呢？至于爱丁顿的论证，在钱德拉塞卡看来似是而非——几乎肯定是错的。

那时候，爱丁顿才是英国天文学的伟人；他的发现几乎都是充满传奇的。天文学家们对太阳和天狼星那样的正常恒星的内部，对它们的大气和它们发出的光的认识，在很大程度上都是靠爱丁顿的发现。所以，学会的会员们和来自全世界的天文学家们，都自然地满怀敬意地听他讲话。显然，如果爱丁顿认为钱德拉塞卡的分析错了，那么它一定是错的。

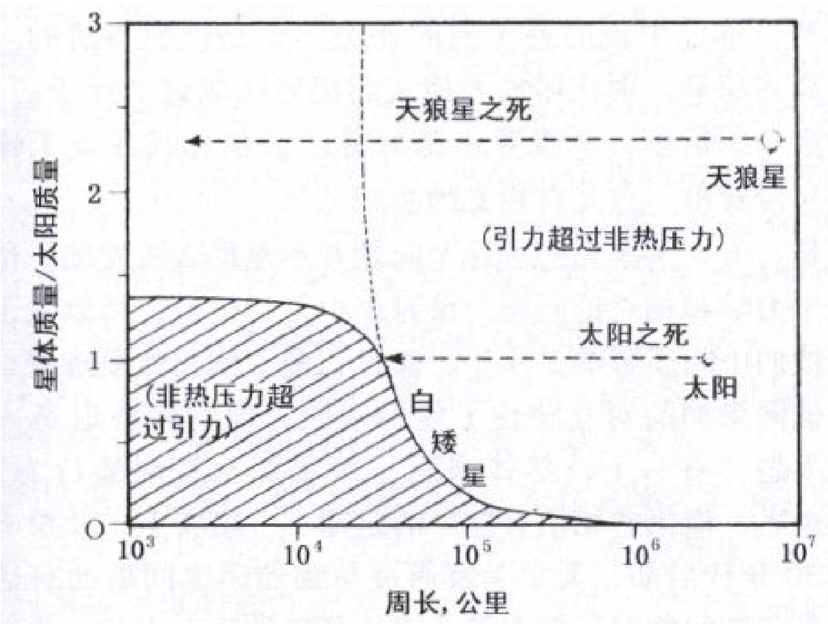



图4.4　像太阳和天狼星（不是天狼B）那样的正常恒星冷却时，一定会发生收缩，沿着图中的质量-周长曲线向左运动。太阳的收缩在到达阴影区边缘（白矮星曲线）时停止，那里的简并压力平衡了引力的挤压。相反，天狼星的收缩不可能这样停下来，因为它永远也不会达到阴影区的边缘。这些结论的不同表示请看卡片4.2。假如照爱丁顿宣布的那样，白矮星物质的压缩阻抗总是5/3，也就是说相对论在高密度下没有将它减小到4/3，那么质量-周长图将具有点虚线的形式，从而天狼星将在这条线上停止收缩。

会后，会员们一个个走到钱德拉塞卡跟前来安慰他。米尔恩告诉他，“我知道爱丁顿是对的，尽管不知道为什么。”

第二天，钱德拉塞卡开始向他的物理学朋友们求助9在给哥本哈

根的罗森菲尔德（Leon Rosenfeld）的信中，他写道："假如爱丁顿是对的，那我最近四个月的工作全都将化为灰烬。爱丁顿能对吗？我非常想知道玻尔的意见。"（尼尔斯·玻尔（Niels Bohr）是量子力学创立者之一，30年代最受尊重的物理学家。）两天后，罗森菲尔德回信了，保证他和玻尔都相信爱丁顿是错的，而钱德拉塞卡是对的。"我想，你的信有点儿令我吃惊，"他写道，"因为没人会那样去追问那些[你用来推导4/3阻抗的]方程，你信中说的爱丁顿的评论完全是模糊不清的。所以你最好高兴起来，别让那些大神父们把你吓成这个样子。"在同一天的后一封信里，罗森菲尔德写道："玻尔和我在爱丁顿的讲话中绝对没看出一点儿有意义的东西。"[10]

但是，天文学家当初对这个问题并不是那么清楚的。他们在这些量子力学和相对论问题上没有多少专门经验，所以爱丁顿的权威在他们中间还影响了几年。而且，爱丁顿仍然坚持着自己的立场。他同黑洞的对立蒙住了他的双眼，他的判断也是一团云雾。他太想"有一个自然律来阻止恒星那么荒谬的行为"了，在他的余生，他仍然相信有这样的定律——事实上，并没有。

到30年代后期，天文学家通过与物理学家同事的对话，认识到了爱丁顿的错误，但由于尊重他早年的巨大成就，他们没有公开这么讲。1939年在巴黎的一次天文学会议的演讲中，爱丁顿又攻击了钱德拉塞卡的结论。在爱丁顿讲话时，钱德拉塞卡给主持人罗素（Henry Norris Russell，来自美国普林斯顿大学的著名天文学家）递了张字条，请允许他答辩。罗素也回了张字条说，"最好不要"，尽管那天早些时候，罗素曾私下告诉钱德拉塞卡，"走出这儿，我们都不相信爱丁顿。"[11]

全世界的主要天文学家最终都——至少背着爱丁顿——接受了钱德拉塞卡的最大白矮星质量，那么，他们愿意承认黑洞能在真实宇宙中存在吗？一点儿也不。假如自然没有提供爱丁顿寻找的那种对抗黑洞的定律，那么它一定会找到别的出路：大概每颗大质量的恒星在老

化和垂死过程中会向星际空间发射足够多的物质，将自身质量减到1.4个太阳以下，从而平安进入白矮星的墓穴。[12]在爱丁顿论战失败后，大多数天文学家都相信这种看法，而且一直坚持到60年代初。

对钱德拉塞卡来说，与爱丁顿的争论给他造成了很大的伤害。大约40年后，他回忆说，“我感到天文学家无一例外地都认为我错了。他们把我看成一心想杀害爱丁顿的堂·吉诃德。你可以想象，当我发现自己在同天文学的巨人论争，而且我的工作完全不被天文学界相信——那对我来说是多么沮丧的经历啊。下一步做什么，我只得自己下决心。我应该在我的余生继续奋斗吗？毕竟那时我才二十四五岁，我想自己还可以做30到40年的科学工作。我根本没有想过拿重复别人做过的事情来当科学创造。对我来说，更好的是改变我的兴趣进入别的什么领域。”[13]

于是，1939年，钱德拉塞卡离开了白矮星和恒星死亡的领域，等四分之一世纪以后他才会回来（第7章）。

那么爱丁顿呢？他对钱德拉塞卡为什么那样无情？对爱丁顿来说，这样的对待似乎一点儿也不过分。在他看来，生命之路，就是吵闹和自由的理性论争。以这种态度对待年轻的钱德拉塞卡，在某种意义上，可能是尊重的表现，也是一个信号，说明他已经接受了钱德拉塞卡，把他作为天文学的建立者中的一员。事实上，从1935年他们第一次对立到1944年爱丁顿去世，爱丁顿对钱德拉塞卡都表现出热情和喜爱，钱德拉塞卡尽管为争论而难过，也还是同样地尊重和爱戴爱丁顿。1

[1] 1982年，钱德拉塞卡发表了纪念A·S·爱丁顿诞辰100周年的演说，全面总结并高度评价了爱丁顿的功绩。这篇演说的中译文收在我们这个“第一推动丛书”系列里《莎士比亚、牛顿和贝多芬——不同的创造模式》一书中（杨建邺、王晓明等译，1997）。——译者

# 第5章　坍缩是必然的

即使我们认为

在所有力中最强的核力

也顶不住引力的挤压

## 茨维基

在三四十年代，弗里兹·茨维基（Fritz　Zwicky）的许多同事都认为他是一个令人头痛的小丑，后代的天文学家回头来看，会认为他是一个有创造力的天才。

“在我1933年认识弗里兹时，他完全相信自己内心有一条通向终极知识的道路，而别人个个都是错的。”[1]福勒这么说，那时他是加州理工学院（加利福尼亚理工学院）的学生，茨维基在那儿教书和研究。从40年代后期就在学院与茨维基同事的格林斯坦（Jesse Greenstein）回忆说，茨维基是“一个喜欢自我宣扬的天才……无疑，他具有与众不同的思想，但是，尽管他不承认，他确实有些粗野，也不是那么能控制自己……他上一门物理课，让谁听课全凭他喜欢，如果他认为谁能够领会他的思想，他就让他来听课……[在学院的物理教授中，]他很孤单，而且不讨学校的喜欢……他常在发表的文章中猛烈攻击别人”。[2]

敦实而自负的茨维基总喜欢斗争——他不懈地宣扬他那通向终极知识的思想路线，夸耀它所带来的启示。30年代，他在一个个演讲、

一篇篇文章中宣传他的中子星概念——这个概念是他茨维基为了解释天文学家观察到的高能现象超新星和宇宙线的起源而创造的。他甚至到全国性的广播节目中去普及他的中子星。[3]但是，仔细研究会发现，他的文章和演说是不能令人信服的，它们没有提供多少能证实他思想的东西。

有一个传说。在茨维基大肆宣扬时，有人问密立根（Robert Millikan，是他将加州理工学院建成为科研院校中的一个强者），为什么把茨维基留在加州，他回答说，也许茨维基的远见中有些是对的。密立根与科学机构的其他一些人不同，他一定感觉到了茨维基直觉的天赋——等到35年之后人们才普遍认识到这一点，那时，实测天文学家在天空中发现了真实的中子星，证实了茨维基关于它们的一些夸大的论断。

在茨维基的论断中，与本书关系最密切的是中子星扮演的恒星遗骸的角色。我们将看到，质量太大的不能成为白矮星的正常恒星死亡后可能成为中子星。假如所有的大质量恒星都以这种方式死亡，那么宇宙就能摆脱那种最奇异的假想的星体遗骸：黑洞。轻恒星死亡后变成白矮星，重恒星变成中子星，看来大自然没有办法生成黑洞了。爱因斯坦和爱丁顿以及他们那个年代的大多数天文学家可以长长地舒一口气。

茨维基是1925年受密立根吸引来到加州理工学院的。密立根希望他在原子和晶体的量子力学结构方面做理论研究，但在20年代后期和30年代初，茨维基越来越多地对天文学感兴趣。在帕萨迪纳工作的人，是很难不为天文学的宇宙着迷的，那儿不仅是加州理工学院的家，也是威尔逊山天文台的家，天文台有着世界最大的直径为2.5米（100英寸）的反射望远镜。




茨维基1931年在加州理工学院的科学家聚会中。照片上还有托尔曼（他是本章后面的重要人物）、密立根和爱因斯坦。[加利福尼亚理工学院档案馆藏。]

1931年，茨维基认识了巴德（Walter Baade），他刚从汉堡和哥廷根来到威尔逊山，是一名优秀的实测天文学家。他们有相同的文化背景：巴德是德国人，茨维基是瑞士人，都讲德语。他们也都钦佩对方的成就。不过，他们的共同之处到此为止。巴德的气质跟茨维基不同，他沉默、高傲，很难接近，消息灵通——而且对同事的怪癖很宽容。从第二年到第二次世界大战期间两人决裂，茨维基常令巴德忍无可忍。“茨维基叫巴德纳粹，但他不是。巴德说，他害怕茨维基会杀他。让他们这对儿住在同一间屋子里是很危险的。”格林斯坦回忆说。4

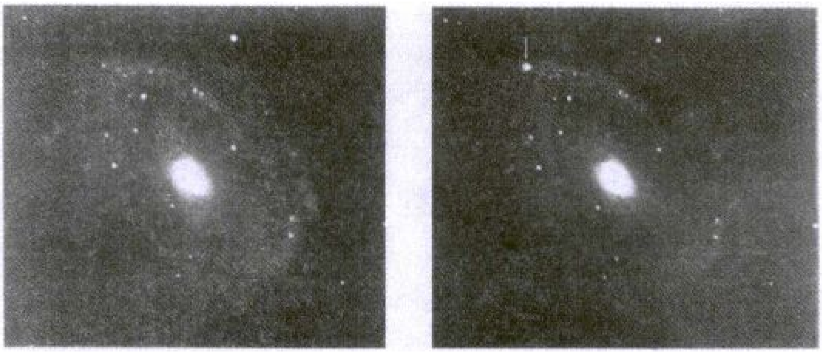
1932和1933年间，巴德和茨维基常在帕萨迪纳用德语热烈讨论叫“新星”的星体，它会突然爆发，比先前亮10 000倍，然后大概过一个月，又慢慢暗淡下来，回到正常状态。巴德凭他渊博的天文学知识知道，有初步的证据表明，除了这些“平常的”新星外，可能还存在不平常的罕见的超级光亮的新星。天文学家起初并没想到这些新星是超亮度的，因为从望远镜看它们表现的亮度跟平常的新星差不多是一样

的。然而，它们处在特殊的星云（闪光的“云”）里。20年代，威尔逊山和别处的观测开始让天文学家相信，那些星云不像我们原先认为的那样只是我们银河系中的气体云，而本来就是一些星系——由近$10^{12}$（万亿）颗恒星形成的巨大集合，远在我们自己的星系之外。我们所看到的这些星系里的罕见新星比我们自己星系里的平常新星遥远得多，为了能够表现出从地球上看到的亮度，它们必然在本质上具有比平常新星强得多的发光能力。

世纪之交以来，天文学家观测到36颗这样的超光亮新星，巴德从已经发表的文献中把能找到的观测数据都收集起来，将这些数据同他得到的有关新星到银河系距离的观测资料相结合，根据结合计算这些超光亮新星发出了多少光。他的结果令人惊讶：在爆发中，超光亮新星的发光本领典型地比我们的太阳大$10^8$（1亿）倍！（现在我们知道，这些距离在30年代被低估了约10倍，[5]相应地，超光亮新星的发光能力应比我们太阳强$10^{10}$（100亿）倍。[1]这在很大程度上也应归功于巴德1952年的工作。）

喜欢极端的茨维基被这些超光亮新星迷住了。他和巴德没完没了地讨论，将它们命名为超新星。他们设想（对的），每颗超新星都是正常恒星爆炸产生的。因为爆炸太热，他们猜想（这回错了），更多的能量是通过紫外线和X射线辐射出去的，而可见光较少。由于紫外线和X射线不能穿透地球大气，所以不可能测量它们包含了多少能量。然而，有可能根据观测到的光谱线和决定在爆炸中超新星热气体行为的物理学定律来估计能量。

将巴德的观测和平常新星的知识与茨维基的理论物理学认识相结合，巴德和茨维基得到结论（错的），来自超新星的紫外线和X射线携带的能量至少是可见光的1万倍或者也许1 000万倍。[6]



后发座内的星系NGC4725。左：1940年5月10日超新星爆发前的照片。右：1941年1月2日超新星爆发中的照片。星系外围白线所指即超新星。现在知道，此星系距地球3000万光年，包含$3\times10^{11}$（3000亿）颗恒星。[加利福尼亚理工学院藏。]

喜欢极端的茨维基立刻认定较大的因子（1 000万）是正确的，而且满怀热情地引用它。

这个（不正确的）1 000万的因子意味着，在超新星最亮的几天内，它将发出巨大的能量：大约是我们的太阳在它100亿年寿命中以热和光的形式辐射出的能量的100倍。这差不多是我们将太阳质量的十分之一转化为纯光能所能获得的能量！

（由于后来几十年的超新星的观测研究——许多工作是茨维基自己做的——我们现在知道，巴德-茨维基超新星能量估计并没偏离太远，但我们也知道，他们的能量计算有严重的缺陷：几乎所有释放的能量实际上都是被称为中微子的粒子带走的，而不是他们认为的X射线和紫外辐射。巴德和茨维基能得到正确答案全靠运气。）

巨大的超新星能量会来自哪里呢？为解释这个问题，茨维基提出了中子星。

茨维基对物理学和天文学的所有分支都感兴趣，他想象自己是一个哲学家。他试图以一种被他后来称为“形态学方式”的方法把他遇到的所有现象都联系起来。1932年，在所有物理学和天文学课题中最受

欢迎的是核物理学，即研究原子核的学问。茨维基从那儿汲取了他的中子星思想的关键要素：中子的概念。

中子在本章和下一章都很重要，我现在暂时离开茨维基和他的中子星，来谈谈中子的发现和它与原子结构的关系。

1926年“新的”量子力学定律建立后（第4章），物理学家在接下来的5年里，用这些量子力学定律去探索微观世界。他们揭开了原子以及那些由原子组成的分子、金属、晶体和白矮星物质等材料的秘密（卡片5.1）。接着，1931年，物理学家将注意力转到了原子的中心和处在中心的原子核。

原子核的性质曾是个大秘密。大多数物理学家认为它是由一组电子和双倍的质子以某种未知的方式束缚在一起而构成的。但是，英国剑桥的卢瑟福（Ernst Rutherford）有不同的假设：原子核由质子和中子构成。那时，已经知道质子是存在的，物理学实验已经研究几十年了，而且还知道它比电子重约2 000倍，带有正电荷。中子还不知道。为了用量子力学定律成功解释原子核，卢瑟福只好假定中子是存在的。成功的解释需要三个条件：（1）中子必须具有大致与质子相同的质量，但不带电荷，（2）每个核必须包含大致相同数目的中子和质子，以及（3）所有中子和质子必须通过一种新类型的力紧密地聚集在小小的核内——这种力既不是电力，也不是引力，自然被称为核力。（现在我们也称它强力）。中子和质子通过幽闭的无规则的高速运动来对抗原子核的约束，这些运动产生简并压力，压力与核力平衡以维持核的稳定和约$10^{-13}$厘米的大小。

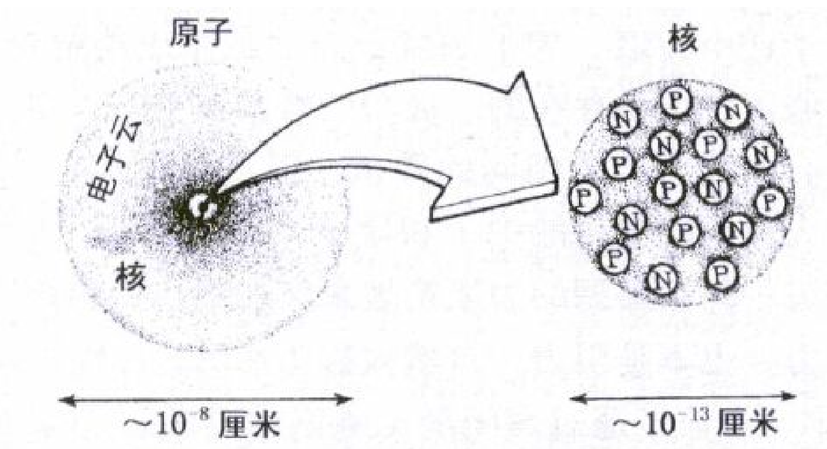
卡片5.1 原子的内部结构

原子由围绕中心的大质量核的电子云组成。电子云的大小约为$10^{-8}$厘米（大概是人的头发直径的百万分之一），处在中心的核要小100 000倍，约为$10^{-13}$厘米，见下面的图。假如把电子云扩到地球大

小，那么原子核有足球场那么大。核尽管小，却比空空的电子云重几千倍。

带负电的电子被带正电的原子核的电作用力吸引而留在电子云里，但它们并不落进核内，原因与白矮星不发生坍缩是一样的：一个叫泡利不相容原理的量子力学定律禁止两个以上的电子在同一时间占据空间的同一个区域（如果两个电子有相反的“自旋”，那是可以的。在第4章，我们忽略了“自旋”这个微妙的东西）。云里的电子于是成对出现在被称为“轨道”的空间。每对电子都不愿局限在它的小格子里，它们将像“幽闭者”那样快速无常地运动，就像在白矮星中的电子一样（第4章）。这些运动提高了“电子简并压力”，抵消了核的电力作用。因此，我们可以认为原子是小小的白矮星，向内吸引电子的不是引力而是电力，而电子简并压力将电子向外推。

下面的右图简单地勾画了刚才讨论的原子核的结构，它是靠核力聚集的质子和中子小集合。




1931年到1932年初，实验物理学家们竞相检验原子核的这种描述。方法是，用高能辐射轰击原子核，设法将卢瑟福假想的中子从核中打出来。1932年，卢瑟福自己的实验小组成员查德威克（James

Chadwick）赢得了竞赛。查德威克的轰击成功了，中子大量出现，正好具有卢瑟福假想的那些性质。世界各地的报纸热热闹闹地宣布了这个发现，茨维基当然也注意了。

中子出现的那年，正是巴德和茨维基在为认识超新星奋斗的一年。在茨维基看来，这个中子正好就是他们需要的东西。[7]他推测，正常星体的核，比如密度为每立方厘米100克，可能会通过坍缩达到像原子核那样的密度，即每立方厘米$10^{14}$（百万亿）克，也许在那个收缩的星体核中的物质会使自己变成中子的“气体”——茨维基称它为“中子星”。假如真是这样，茨维基算出（这回对了），收缩核的强大引力会紧紧地将它束缚起来，这不仅会减小它的周长，还会减少它的质量。星体核现在的质量将比它坍缩前轻10%。那10%的质量跑到哪儿去了呢？茨维基猜想，它们变成了爆发的能量（这回他又对了。见图5.1和卡片5.2）。

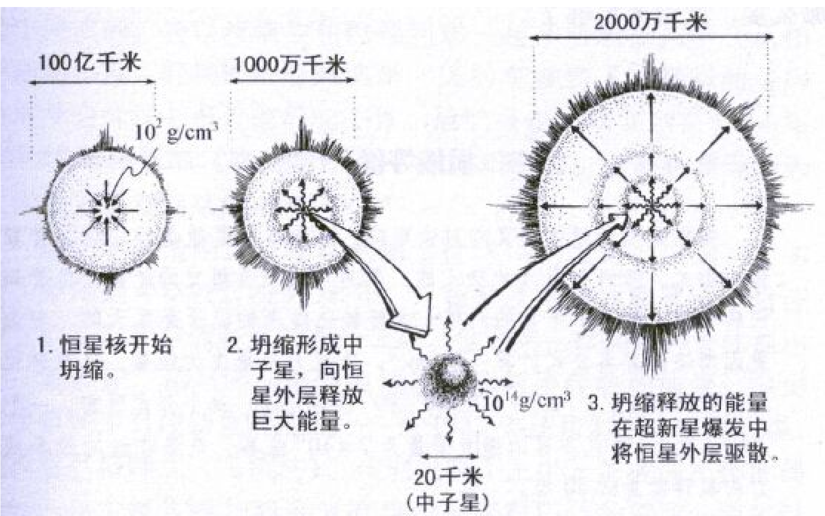



图5.1　茨维基触发超新星爆发的假设：超新星爆发的能量来自星体从正常密度核坍缩成中子星的过程。

茨维基相信（对了），假如星体收缩核的质量与太阳的质量相同，那么当核转变成中子星时，那10%的质量将转变为爆发的能量，

约$10^{46}$焦耳，接近茨维基所想的超新星需要的动力。爆发的能量能将星体外层加热到很高的温度，然后将它们吹散在星际空间（图5.1）；星体爆炸时，因高温而发光，那正是他和巴德所认定的超新星的行为。

茨维基不知道什么东西能引起星核坍缩而使它转变为中子星，也不知道核在坍缩时有什么行为，所以他没法估计坍缩会持续多长时间（是慢慢收缩，还是迅速坍缩？）。（60年代，我们终于发现了所有的细节，原来坍缩是很剧烈的；强大的引力在不到10秒钟内就使星体从地球那么大坍缩到100公里的周长。）茨维基也没有完全认识核收缩的能量如何能产生超新星爆发，或者说，他不理解，为什么爆炸的碎屑在几天里会那么耀眼，而且还能亮几个月，而不是几秒钟、几个小时或者几年？然而，他知道——或者说他认为他知道——在中子星形成时放出的能量正好是那么多，这对他足够了。

卡片5.2　质能等价

根据爱因斯坦的狭义相对论定律，质量不过是能量的一种非常紧致的形式。任何质量（包括人的）都可能转化为爆发的能量，尽管如何转化并不是一个平常的问题。这样转化而来的能量是巨大的，它由爱因斯坦的著名公式计算：$E=mc^2$，这里E是爆发的能量，M是转化为能量的质量，$c=2.99792\times10^8$米/秒是光速。这个公式预言，一个正常人的75千克质量的爆发能量为$7\times10^{18}$焦耳，是爆炸过的最具威力的氢弹能量的30倍。

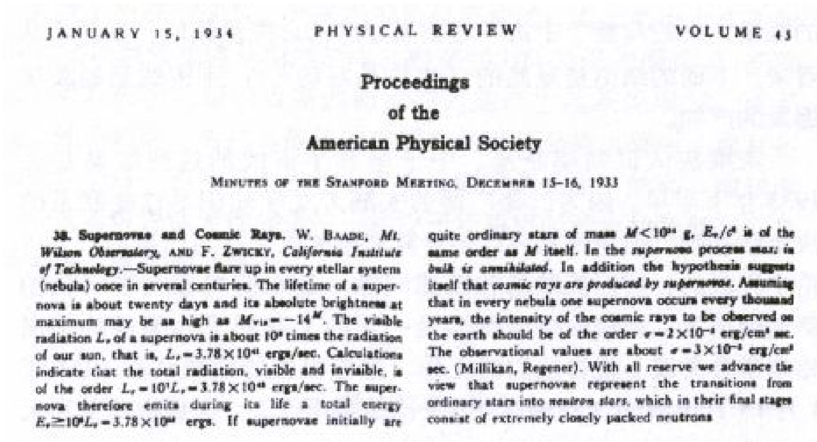
质量转化为热量或爆发的动能，是茨维基解释超新星（图5.1）、解释维持太阳热量的核燃烧（本章后面）和核爆炸（下一章）的基础。

茨维基并不满足于解释超新星，他还想解释宇宙间的一切事物。1932～1933年，在所有还没解释的事物中，在加州理工学院最受注

意的是宇宙线——来自空间的轰击地球的高速粒子。学院的密立根是宇宙线研究的世界领袖，他为它们命了名；安德森发现有些宇宙线是由反物质构成的。[2]爱走极端的茨维基想让自己相信，大多数宇宙线来自我们太阳系以外（这是对的），而且大多数还来自我们银河系之外——实际上，来自宇宙最遥远的地方（错了）——然后，他相信（基本是对的），所有宇宙线携带的能量，大约与整个宇宙间超新星释放的总能量相同。在茨维基看来，下面的结论是显然的（也许是对的[3]）：宇宙线是超新星爆发的产物。

茨维基认识到超新星、中子星和宇宙线的这些联系是在1933年下半年。因为巴德广博的实测天文学知识是这些联系的决定性基础，而茨维基的许多计算和推测也是在与巴德的口头讨论中完成的，所以茨维基和巴德同意一起在斯坦福大学（从帕萨迪纳出发，轻轻松松沿海滨坐一天的车就到了）举行的美国物理学会会议上报告他们的工作。他们报告的摘要刊在1934年1月15日出版的《物理学评论》上，如图5.2。这是物理学和天文学史上最富远见的文献之一。[8]

他们明确地断言，存在超新星这样一类独特的天体——不过，要等到4年以后的1938年，巴德和茨维基才能拿出严格证明它们不同于普通新星的充分证据。他们第一次为这种天体提出了“超新星”的名字；正确地估计了超新星释放的能量；提出了宇宙线来自超新星的建议——在1933年还是似乎合理的假说，但没有严格确立（见脚注）。他们发明了由中子形成的星体的概念——这个概念到1939年才在理论上获得广泛的赞同，而实际观测确认要等到1968年。他们为这个概念起了中子星的名字。他们还“完全保留”（这大概是谨慎的巴德插进的一个短语）地提出，超新星是正常星向中子星转化的产物——到60年代初，人们才看到这个建议在理论上是可能的，但等到60年代末在古代超新星爆发的气体内发现脉冲星（旋转的磁性的中子星）后，它才得到观测的证明。



JANUARY 15, 1934 PHYSICAL REVIEW VOLUME 45

**Proceedings of the American Physical Society**

MINUTES OF THE STANFORD MEETING, DECEMBER 15–16, 1933

**38. Supernovae and Cosmic Rays.** W. BAADE, *Mt. Wilson Observatory*, AND F. ZWICKY, *California Institute of Technology*.—Supernovae flare up in every stellar system (nebula) once in several centuries. The lifetime of a supernova is about twenty days and its absolute brightness at maximum may be as high as $M_{vis} = -14^M$. The visible radiation $L_v$ of a supernova is about $10^8$ times the radiation of our sun, that is, $L_v = 3.78 \times 10^{41}$ ergs/sec. Calculations indicate that the total radiation, visible and invisible, is of the order $L_\tau = 10^7 L_v = 3.78 \times 10^{48}$ ergs/sec. The supernova therefore emits during its life a total energy $E_\tau \geq 10^5 L_\tau = 3.78 \times 10^{53}$ ergs. If supernovae initially are quite ordinary stars of mass $M < 10^{34}$ g, $E_\tau/c^2$ is of the same order as $M$ itself. In the *supernova process mass in bulk is annihilated*. In addition the hypothesis suggests itself that *cosmic rays are produced by supernovae*. Assuming that in every nebula one supernova occurs every thousand years, the intensity of the cosmic rays to be observed on the earth should be of the order $\sigma = 2 \times 10^{-3}$ erg/cm² sec. The observational values are about $\sigma = 3 \times 10^{-3}$ erg/cm² sec. (Millikan, Regener). With all reserve we advance the view that supernovae represent the transitions from ordinary stars into *neutron stars*, which in their final stages consist of extremely closely packed neutrons.

图5.2　巴德和茨维基1933年12月在斯坦福大学关于超新星、中子星和宇宙线的报告的摘要。

30年代的天文学家热烈响应巴德-茨维基超新星概念，但对茨维基关于中子星和宇宙线的思想却不大看得起。普遍认为它“太假了”；可能还有人说它“建立在不可靠的计算基础上”，说得很对。茨维基的文章和谈话没有为他的思想提供更充分的证明线索。事实上，我在仔细研究茨维基那个年代的文章后发现，他没有很好地理解物理学定律，所以没能证明他的想法。本章后面我还会回来谈这一点。

回溯起来，有些科学概念本来是很显然的，所以我们会奇怪，当时竟没有人马上认识到它们。中子星与黑洞的联系就是这样的例子。茨维基在1933年就可能发现这个联系，但他没有；6年以后，这种联系才初步建立起来，而在20多年以后它才确定下来。这段令物理学家想起来就脸红的曲折经历，是本章后面的主题。

物理学家是如何认识到中子星-黑洞联系的呢？为讲好这个故事，先应该知道关于这个联系的一些事情，下面我们就来看看。

恒星死后的命运如何？第4章揭示了部分答案，画在图5.3的右边（与图4.4是一样的）。那个答案依赖于恒星的质量是小于还是大于1.4个太阳（钱德拉塞卡极限质量）。

假如恒星质量小于钱德拉塞卡极限，比如太阳自己，它在生命的尽头将走上图5.3中写着“太阳之死”的路径。它一面向空间辐射光，一面慢慢冷却，失去（热产生的）热压力。因为压力减小了，它不再能够抵抗引力向内的挤压。引力迫使它收缩。收缩时，它在图5.3中向着左端移动到更小的周长，而总是处在图中相同的高度上，因为它的质量不会改变。（注意，图中质量向上、周长向右增加。）收缩时，内部的电子被挤到越来越小的格子里。最后，电子以它强大的简并压力顶住了星体的收缩。简并压力对抗着星体向内的引力挤压，迫使它在图5.3的白区和阴影区的边界曲线（白矮星曲线）上安静下来，走进白矮星的墓穴。如果星体还要收缩，（也就是从白矮星曲线向左走进阴影区），电子简并压力会变得更强，使星体膨胀又回到白矮星曲线。如果星体还要膨胀到白区，电子简并压力将减弱，引力又使它收缩，回到白矮星曲线。这样，星体除了永远停留在白矮星曲线上，没有别的选择。引力和压力在这里完全平衡，它将慢慢冷却而变成一颗黑矮星——一个地球大小却具有太阳质量的冰冷而黑暗的天体。

假如恒星质量超过了钱德拉塞卡1.4个太阳质量的极限，比如天狼星，它在生命的尽头会走上“天狼星之死”的道路。当它发出辐射，冷却收缩，沿着那条道路向左走向越来越小的周长时，它的电子也被挤到越来越小的格子里；这些电子以不断增大的简并压力反抗着，然而反抗是徒劳的。因为恒星质量太大，引力足以压倒一切电子的反抗。电子不可能产生足够的简并压力来对抗恒星的引力，[4]恒星必然像爱丁顿说的那样，“继续地辐射下去，收缩下去，我想，它会一直辐射收缩到几公里的半径，那时引力会变得很强大，足以平息这些辐射，而恒星也最终找到了安宁。”

如果不是中子星，恒星的命运可能会是这样的。假如茨维基是对的，中子星能够存在，那么它们一定会像白矮星，不过内部压力是中子而不是电子产生的。这意味着在图5.3中应该有一条类似于白矮星曲线的中子星曲线，但周长（标在水平轴上）大约是几百公里，而不

是几万公里。在中子星曲线上，中子压力完全与引力相平衡，所以中子星会永远留在那儿。

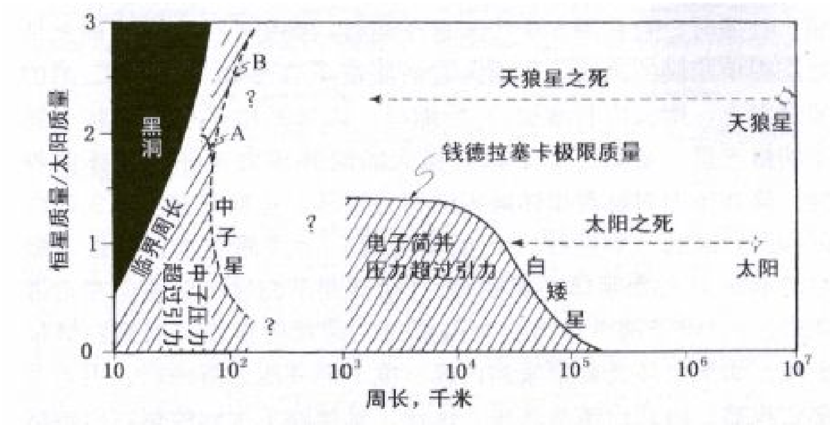



图5.3　质量大于1.4个太阳质量的钱德拉塞卡极限的恒星的最终命运，依赖于中子星的质量能有多大。假如中子星质量可以任意大（曲线B），那么像天狼星那样的恒星死亡后只能坍缩成中子星，不能形成黑洞；假如中子星存在质量上限（如曲线A），那么大质量的死亡恒星既不能成为白矮星，也不能成为中子星；如果没有别的归宿，它们就只能成为黑洞。

假定中子星曲线在图5.3中向上延伸到更大的质量，就是说，假定它像图中的曲线B那样，那么，天狼星死亡时不会生成黑洞。相反，它会收缩到中子星曲线，然后就不再收缩了。如果还要收缩（就是说，向中子星曲线的左端移动，进入阴影区），它内部的中子将反抗挤压，产生巨大的压力（部分是因为简并，即“幽闭”，部分是核力）；压力很大，足以超过引力从而使星体向外扩张，回到中子星曲线。如果星体想扩张到白区，中子的压力将降低，又让引力占上风，将星体挤压回来。这样，天狼星没有别的选择，只好停在中子星曲线上，永远留在那儿，慢慢冷却下来，成为一颗致密、冰冷而黑暗的中子星。

假定中子星曲线不向上延伸到更大的质量，而是像假想曲线A那样弯曲。这意味着任意中子星都存在一个极大质量，类似于白矮星的1.4个太阳质量的钱德拉塞卡极限。对中子星来说，也像白矮星的情

形一样，极大质量的存在预示着一个极其重要的事实：在质量超过极大值的恒星内，引力将完全超越中子的压力。于是，当这么大质量的恒星死亡时，它要么放出足够多的质量而低于极大值，要么在引力作用下无情地收缩下去，通过中子星曲线，然后——假如除了白矮星、中子星和黑洞以外，没有其他可能的星体归宿——它会一直收缩下去，形成一个黑洞。

因此，中心问题，也就是关系着大质量恒星的最终命运的问题是，一个中子星能有多大质量，如果可以很大，大于任何正常星体，那么黑洞就不可能在真实宇宙中形成。如果中子星有一个极大的可能质量，而那个极大质量又不是太大，那么黑洞是会形成的——除非还有什么别的在30年代未曾想到过的恒星归宿。

现在看来，这条推理路线那么明显，但茨维基没走上来，钱德拉塞卡没走上来，爱丁顿也没走上来，真是令人奇怪。不过，就算茨维基走上来了，也不会走得太远，他对核物理和相对论了解得太少，不可能发现物理学定律是否为中子星安排了质量极限。然而，在加州理工学院真有两位懂物理的人，他们能推导中子星的质量：一个是托尔曼（Richard Chace Tolman），他从化学家变成物理学家，写过一本题为《相对论、热力学和宇宙学》的经典教科书；另一个是奥本海默（J.Robert Oppenheimer），他后来将领导美国发展原子弹。

但是，托尔曼和奥本海默对茨维基的中子星一点儿兴趣也没有，直到1938年他们才注意到它，那年，中子星的思想发表了（名字稍有不同，叫中子核），作者是另一个人，他不像茨维基那样不太讨人喜欢，而是大家尊敬的莫斯科的朗道（Lev Davidovich Landau）.

[1] 地球接收到的光总量反比于到超新星的距离的平方，所以距离的10倍误差意味着巴德估计的光的总输出量有100倍的误差。

[2] 反物质的名字源于这样的事实：当物质粒子遇到反物质粒子时，它们会彼此湮灭。

[3] 现在看来，宇宙线有许多不同的来源。但我们还不知道哪种方式产生的宇宙线最多，不过，很有可能是，在超新星爆发很久以后，残留气体云的激波将粒子加速到很高的速度。如果确实如此，那么在间接意义上说，茨维基是对的。

[4] 理由在卡片4.2中解释过。

# 朗道

朗道关于中子核的文章实际上是一篇求助的呼唤。[9]那时，斯大林的清洗运动席卷整个苏联，朗道很危险。他希望以他的中子核思想在报纸上激起大波，这样可能会使他免遭逮捕和死亡，但关于这点，托尔曼和奥本海默一无所知。

朗道的危险来自他过去与西方科学家的接触。

俄国革命不久，科学成了新共产党领导人特别关注的目标。列宁促使在1919年的布尔什维克党第八次会议上通过一项决议，免除了对科学家的意识形态纯洁性的要求：“工业和经济发展的问题需要立刻广泛地发挥我们从资本主义那里得来的科学和技术专家们的作用，尽管他们难免沾染些资产阶级的思想和作风。”特别令苏维埃科学领导者们关心的是苏联理论物理学的可怜状况，所以，托共产党和政府的福，苏联最有才华和希望的年轻理论家们来到列宁格勒（圣彼德堡）读了几年研究生，完成相当于博士学位的课程后，被送到西欧作一两年的博士后研究。

为什么要读博士后呢？因为到20年代，物理学已经太复杂了，博士水平的培养不能满足精通的需要。为了在全世界促进更高的培养，一个博士后奖学金体系就建立起来了，主要靠洛克菲勒基金的资助（来自资本主义石油企业的好处）。任何人，即使是热情的苏联马克思主义者，都能竞争奖学金，获奖者就被称为“博士后研究生”或简称“博士后”。

为什么到西欧去读博士后呢？因为在20年代，西欧是理论物理学的圣地，几乎是每一个世界知名的理论物理学家的故乡。苏联的领导者们为了显示他们的恩赐，要把西方的理论物理灌输到苏联，没有别的选择，只好将他们年轻的理论家送到那儿去培养，也顾不上精神污

染的危险了。

在经历过到列宁格勒，然后去西欧，然后回苏联的年轻苏维埃理论家中，朗道在物理学界是最有影响的。他1908年出生在一个小康的犹太家庭（父亲是里海之滨巴库的石油工程师），他16岁进列宁格勒大学，19岁本科毕业。在列宁格勒技术物理学院只读了两年的研究生，他就完成了相当于博士学位的学习，然后来到西欧。在1929～1930年的18个月里，他走遍了瑞士、德国、丹麦、英国、比利时和荷兰的大理论物理学中心。

德国出生的皮尔斯（Rudolph Peierls）是朗道在苏黎世的博士后同学，他后来回忆说，“我还清晰记得朗道1929年在苏黎世出现在泡利的系里时，给我们留下的深刻印象。……没过多久就他发现他对现代物理学的深刻认识和他解决基础问题的技巧。他很少详细阅读理论物理学的论文，只是大概看看，问题是否有趣，如果有趣，作者的方法是什么。然后他开始自己计算，如果答案和作者的一致，他就赞同这篇文章。”[10]皮尔斯和朗道成了最要好的朋友。

朗道高而瘦，对别人、对自己都很严厉。他很失望自己晚生了几年。他认为，物理学的黄金年代是1925～1927年，那时德布罗意、薛定谔、海森伯、玻尔等人正在开创新的量子力学。如果生得早些，他朗道也能加入其间了。“所有的漂亮女孩儿都被抢走跟人结婚了，所有好的物理问题都被解决了。我实在不喜欢剩下的那些东西。”[11]1929年，他在柏林曾这样失望地抱怨。但是，实际上，量子力学和相对论的结果的探索才刚刚开始，那些结果也会带来惊奇：原子核的结构、核能、黑洞和它们的蒸发、超流、超导、晶体管、激光以及磁共振图像等，这只不过是几个例子。朗道虽然悲观，但他将成为探寻这些结果的核心人物。



左：20年代中叶在列宁格勒读书的朗道。右：在列宁格勒读书期间（约1927年），朗道和物理系同学盖莫夫和Yevgenia Kanegiesser在开玩笑。其实，朗道从来不玩任何乐器。[左：美国物理学联合会（AIP）Emilio Segrè图像档案馆，Margarethe Bohr藏；右：国会图书馆藏。]

朗道是一个热情的马克思主义者和爱国者，1931年一回到列宁格勒，他就决心集中精力向苏联输入现代理论物理学，在后面的章节我们会看到，他取得了巨大的成功。

朗道回国不久，斯大林的铁幕降下来，再去西方几乎不可能了。据朗道在列宁格勒的同学盖莫夫（George Gamow）后来回忆，“俄国科学现在成了与资本主义世界斗争的武器。跟希特勒将科学和艺术分为犹太的和亚利安的一样，斯大林发明了资产阶级科学和无产阶级科学的名词。俄国科学家同资本主义国家科学家‘友好’……成了一种犯罪。”[12]

政治气候从恶劣走向恐怖。1936年，在强迫农业合作化过程中杀害了六七百万农民和富农（土地所有者）的斯大林，开始了对全国政界和知识界领导人长达7年的清洗，现在称这场清洗为大恐怖。清洗

处决了几乎所有原列宁的政治局成员。苏维埃军队的最高指挥官们，71名共产党中央委员会委员中的50名，大多数的驻外使节以及非俄罗斯联邦的总理和高级官员，都在清洗中被处决或者被迫消失，永远不再出现。据保守估计，大约700万人被抓进监狱，250万人死亡——其中一丰是知识分子，包括大批科学家和一些研究群体。苏联的生物学、遗传学和农业科学被毁了。[13]

1937年下半年，朗道（他现在是莫斯科理论物理学研究的领导者）感到清洗的恶浪正向他逼近。他在惶恐中寻求保护，一个可能的办法是，让公众都来注意他这个知名的科学家，于是他在他的科学思想中找一个可能在西方和东方都激起巨澜的东西。他选的是从30年代就开始思考的一个思想：像太阳那作的“正常”恒星，可能会在中心拥有中子星——朗道称它们是中子核。

朗道是这样想的：太阳和其他正常恒星通过（热产生的）热压力来抵抗引力而维持自身。太阳向空间辐射热和光时，必然要冷却、收缩，在大约3000万年的时间里慢慢死亡——除非它有什么办法补充失去的热量。在20年代和30年代，有令人信服的证据表明，地球在10亿年或更长的时期内保持着大致相同的温度，所以太阳也一定通过某种方式补充它的热量。爱丁顿等人已经在20年代提出（对的），新的热量可能来自核反应，在反应中，一类原子核转变成另一类原子核——现在我们称它为核燃烧或核聚变，[14]见卡片5.3。不过，到1937年，核燃烧的细节还了解太少，物理学家不知道它是否能够这样。朗道的中子核提供了一个很诱人的可能。

茨维基曾想象，在正常星坍缩成中子星时释放的能量为超新星提供了动力，同样，朗道也能想象，太阳或其他正常恒星也是通过它们的原子一个个被中子核捕获时释放的能量来补充动力的（图5.4）。

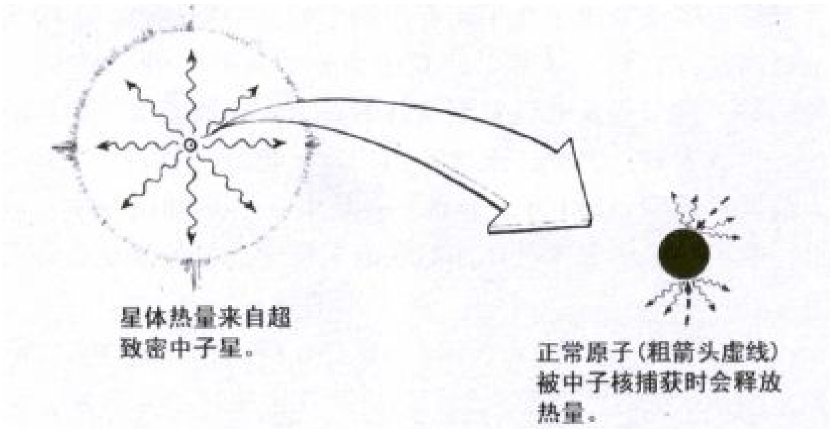
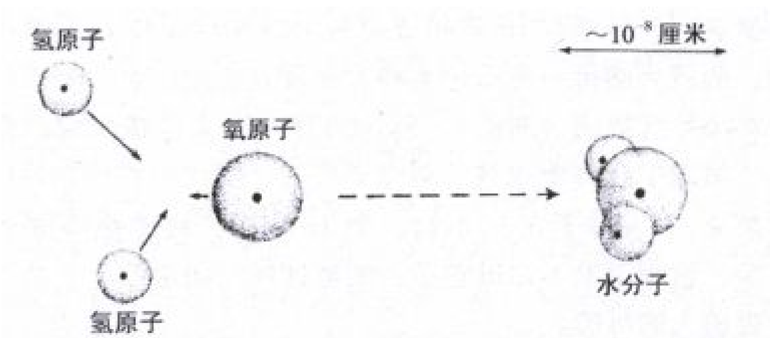



图5.4　朗道关于维持正常恒星热量的能源的猜想

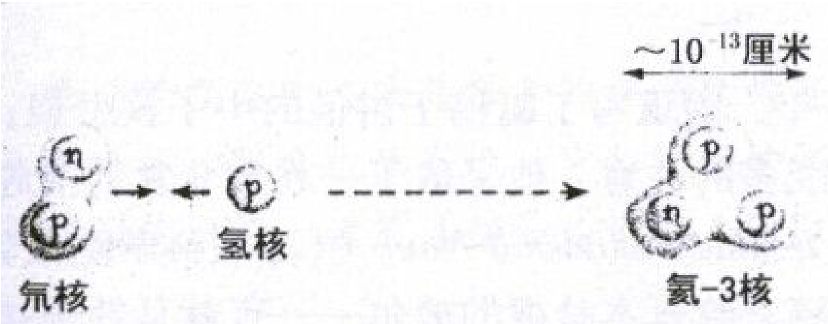
卡片5.3　核燃烧（聚变）与普通燃烧的对比

普通的燃烧是一种化学反应。在化学反应中，原子结合成分子，在分子中共享彼此的电子云，电子云将分子维系在一起。核燃烧是一种核反应。在核反应中，原子核聚在一起（核聚变）形成质量更大的核；核力维系着这个更大质量的核。




上图是普通燃烧的一个例子：氢燃烧生成水（一种通过燃烧提供爆发性动力的形式，常用来作为火箭的动力，送卫星上天）。两个氢原子和一个氧原子结合形成一个水分子。在水分子中，氢氧原子分享彼此的电子云，但它们的核还是独立的。

下图是核燃烧的例子：氘（“重氢”）核与普通氢核聚合形成氦-3核，我们现在知道，这种核聚变反应是太阳和其他恒星的动力，也是氢弹的动力（第6章），氘核包含一个中子和一个质子，由核力束缚在一起；氢核只含有一个质子；聚变生成的氦-3核包含一个中子和两个质子。




原子被中子核捕获很像一块石头从很高的地方落到水泥板上：引力将石块拉下来，使它获得很高的速度，当它打在水泥板上时，巨大的动能（运动的能量）会使它摔得粉碎。同样，朗道认为，中子核上的引力能使塌陷的原子达到很高的速度。当这些原子落进核时，会碎裂开来，动能（相当于总质量的10%）将转化为热量。在这样的图景中，太阳的最终热量来源是中子核的强大引力；而对于茨维基的超新星，核的引力在塌陷原子的质量转化为热量的过程中只有10%的作用。

与中子核捕获原子（图5.4）相比，核燃料的燃烧（卡片5.3）只能将百分之零点几的燃料质量转化为热量。换句话说，爱丁顿的热源（核能）大约比朗道的热源（引力能）小30倍。[1]

朗道在1931年实际上发现了他的中子核思想的更原始的形式。但是，那时还没有发现中子，原子核也还是个谜，所以，在他1931年的模型里，核捕获原子释放能量是通过一个完全假想的过程，而过程的

基础（错的）在于怀疑量子力学定律在原子核问题上可能会失败。[15] 现在中子发现5年了，原子核的性质也开始为人们所认识，朗道可以让他的思想更精确更服人了。把它向世界提出来，在大众中激起波澜，他也许能够躲过斯大林大清洗的浪头。

1937年下半年，朗道写了篇稿子讲他的中子核思想；[16]为保证它能受到尽可能多的注意，他采取了一系列非常的措施：他把稿子交给苏联的Doklady Akademii Nauk（《苏联科学院报告》）发表，同时将英文稿寄给哥本哈根的玻尔——也就是钱德拉塞卡受爱丁顿攻击时曾求助过的那位西方著名物理学家（玻尔作为苏联科学院荣誉院士，即使在大恐怖年代，也多少还能为苏联当局所接受）。同稿子一起，朗道还给玻尔写了下面的信：

亲爱的玻尔先生！

我随信给您寄来一篇我写的关于星体能量的文章。如果您认为它还有点儿物理意义，请您把它交给《自然》。如果您不嫌太麻烦，我会乐意听听您对这个工作的意见。

万分感谢。

您的，L·朗道

1937年11月5日，莫斯科

（《自然》是英国科学杂志，及时发表一切科学领域的发现，也是世界各类科学期刊中发行量最大的之一。）

朗道有个身居高位的朋友——他一听到玻尔认可了他的文章并交给了《自然》，就让《消息报》（苏联最有影响的两家报纸之一，是由苏联政府管理并代表政府的报纸）编辑部给玻尔发了一封电报。电报是1937年11月16日发出的：

请告知您对朗道教授工作的意见。电告您的简单结论。

《消息报》编辑部

玻尔显然对这个要求感到疑惑和忧虑，当天就从哥本哈根回了信：

朗道教授关于大质量恒星的中子核的新思想是很杰出而大有希望的。我会很高兴地对它和朗道的多个其他研究发表一个简短的评价。请更详细地告诉我，为什么要我的意见。

玻尔

《消息报》编辑部回答说，他们想把玻尔的意见在报上发表。11月23日，他们真那么做了，那是一篇讲述朗道的思想并给予高度赞扬的文章：

朗道教授的工作在苏联科学家中间激起了极大的兴趣，他大胆的思想为天体物理学中最重要的过程之一带来了新的生命。有充分的理由认为，朗道的新假说将被证明是正确的，它将为天体物理学中的一大堆尚未解决的问题带来答案。……尼尔斯·玻尔对这位苏联科学家[朗道]的工作作了极高的补充评价，说“L·朗道的新思想是很杰出而大有希望的。”[17]

这场运动还是没能挽救朗道。1938年4月28日清晨，他寓所的门响了，未婚妻科娜在门里震惊地看到他被一辆黑色官方轿车带走了。许多人遭遇的命运现在也降临到了朗道的头上。

轿车将朗道带到莫斯科最臭名昭著的政治监狱，布提斯卡雅。人家告诉他，他的德国间谍活动已经暴露了，他得为此付出代价。滑稽的罪名是无关紧要的（朗道，一个犹太人，热情的马克思主义者，为纳粹德国做间谍？），实际上那时几乎所有的罪名都是可笑的。在斯

大林的俄国，很少有人知道被抓进监狱的真正原因——不过，关于朗道的案子，在最近公开的克格勃文件中可以找到一点儿线索：[18]他在与同事的谈话中，批评过共产党和苏联政府的科研组织方式和大恐怖引发的1936～1937年的大逮捕。这些批评被认为是“反苏联行为”，很容易让人进监狱。

朗道很幸运，他在狱中只呆了一年就活着出来了——这是少有的。他在1939年4月被释放，卡皮查（Pyotr Kapitsa）救了他。卡皮查是30年代苏联最著名的实验物理学家，他直接向莫洛托夫和斯大林说明了理由：朗道，而且在苏联所有的理论物理学家中，只有朗道有能力解开超流出现之谜。[2]（超流是在卡皮查实验室发现的，英国剑桥的阿伦（J.F.Allen）和米斯纳（A.D.Misener）也独立发现了。如果苏联科学家能解释这个现象，就能双倍地向世界证明苏联科学的能力。）[19]

朗道从监狱出来，憔悴不堪，病情严重。终于，他在身心恢复后，用量子力学定律解开r超流之谜，赢得了诺贝尔奖。但他的精神崩溃了，再也经受不住来自政治集团的哪怕一点点精神压力。

[1] 认为核力远比引力更有威力的人看来，这是很奇怪的。只有几个原子或原子核时，核力确实有很强的动力。不过，当有几个太阳或更多的原子（$10^{57}$个）时，所有原子聚在一起的引力将在动力上远远超过它们的核力。在本章后面我们会看到，这个简单的事实最终会保证，当大质量恒星死亡时，巨大的引力将超过原子核的排斥而将它们挤压成一个黑洞。

[2] 超流是某些流体冷却到绝对零度以上几度（也就是冷却到约-270℃）时出现的一种完全没有粘滞力（内摩擦力）的状态。

# 奥本海默

在加利福尼亚，奥本海默照习惯仔细读了朗道发表的每一篇文章。于是，朗道发表在1938年2月19日《自然》杂志上的关于中子核的文章立刻就引起了他的注意。茨维基提出的中子星为超新星提供能源的思想，在奥本海默看来，是容易破碎的幻想，而朗道提出的为正常恒星提供能源的中子核则值得认真考虑。也许，太阳真有这样的核？奥本海默发誓要把它弄清楚。

奥本海默的研究风格与我们迄今为止在本书遇到的任何一个人都不同。巴德和茨维基一起工作，是平等的伙伴，两人的才能和知识互为补充；钱德拉塞卡和爱因斯坦喜欢一个人做研究；而奥本海默热情洋溢，身边总簇拥着一大群学生。爱因斯坦曾为教学感到痛苦，而奥本海默却是在讲课中成长起来的。

跟朗道一样，奥本海默也到过西欧理论物理学圣地学习；跟朗道一样，奥本海默一回家，就着手把从欧洲学来的理论物理学传给他的祖国。

回到美国时，奥本海默赢得了巨大的荣誉，包括哈佛和加州理工在内的十所美国大学和两所欧洲大学都为他提供了教授的职位。这些邀请中，有一个来自在伯克利的加利福尼亚大学，那儿根本没有理论物理。奥本海默后来回忆，“我访问了伯克利，我想我应该去那儿，因为那儿是荒漠。”他在伯克利能开创一些完全属于他个人的东西。不过，奥本海默同时接受了伯克利和加州理工的邀请，因为他怕在学术上孤独。秋天和冬天他在伯克利，春天在加州理工学院。“我保持着与加州理工的联系……如果我偏离基础太远了，那儿能让我走回来；我还能从那儿学到一些在发表的文献里可能反映得不够充分的东西。”

奥本海默刚当老师时，对学生太严厉，太没有耐心，太傲慢。他不了解学生知道多少，也不愿让自己适应他们的水平。1930年春，他在加州理工学院的头一课真是讲绝了——论证有力，语言优美，见解深远。课讲完了，人都走了，托尔曼（这位从化学家来的物理学家现在是他的亲密朋友）跟在身后，让他回到现实："好的，罗伯特，"他说，"讲得太好了，但我一个该死的词儿也没听懂。"[20]

然而，奥本海默很快就会讲课了。一年中，研究生和博士后从美国各地聚到伯克利来跟他学物理。几年内，他就让伯克利成为在美国理论物理学博士后看来比欧洲更有吸引力的地方。

奥本海默的一个博士后塞伯（Robert Serber）后来讲述了跟他工作的感受："奥比（伯克利的学生都这么叫他）反应快，很急躁，讲话刻薄，刚当老师的时候，听说他对学生很严厉，但经过5年后，他成熟了（愿他早年的学生也相信这一点）。他的[量子力学]课能激发人的灵感，也是一个教育成果。他让学生感到物理学的逻辑结构那么美妙，物理学的发展那么激动人心。这门课几乎每个人都听过不止一次，奥比有时劝学生不要来听第三次、第四次，但劝不住……"

"奥比同他的研究生们的合作方式也是前所未有的。他的小组由8～10名研究生和大约六七名博士后组成。每天他在办公室同小组见一面。会见前，组员们陆续进来，在桌旁和墙边坐下。奥比走进来，逐个地同他们讨论研究问题的状况，别的人在旁听着，发表意见。所有的人都面对着广泛的问题。奥本海默对什么事情都感兴趣，一个题目接着一个题目地来，彼此依存。下午，他们可能讨论电动力学、宇宙线、天体物理和核物理。"

每年春天，奥本海默把书和论文塞进他的敞篷车里，后面坐上几个学生，开到帕萨迪纳。塞伯说，"放弃我们在伯克利的房子和公寓是无所谓的，我们相信在帕萨迪纳可以找到一个月租金25美元的花园房子。"[21]

针对每个令他感兴趣的问题，奥本海默都会选一个学生或博士后去研究它的细节。对朗道的问题，中子核能否维持太阳的热量，他选择了塞伯。

奥本海默和塞伯很快发现，假如太阳的中心有中子核，假如核的质量占太阳质量的大部分，那么核的强大引力将紧紧地抓住太阳的外层物质，使太阳周长远远小于实际的长度。于是，朗道的中子核思想只有在核的质量远远小于太阳质量时才能成立。

“中子核的质量能有多小？”奥本海默和塞伯被迫这样问自己。“中子核可能的极小值是什么？”注意，这个问题是与黑洞存在的决定性问题相反的：为知道黑洞是否能够形成，我们需要知道一颗中子星的最大可能质量（上面图5.3）。奥本海默还一点儿也没看到极大质量问题的重要性，但他现在知道，对朗道的思想来说，中子核的极小质量是关键的。



塞伯（左）和奥本海默（右）在讨论物理，约1942年。[美国新闻局提供。]

朗道在文章里也知道中子核极小质量的重要，还用物理学定律估算过。奥本海默和塞伯仔细审查了朗道的估计。他们看到，朗道恰当地考虑了核内部和附近引力的吸引，这是对的；他考虑了核的中子的简并压力（当中子被挤压到一个小空间内时由中子的幽闭运动产生的压力），也是对的；但是，他没有考虑中子彼此之间的核力，这就错了。那种力当时也还没有完全明白，不过对奥本海默和塞伯来说，已经足以得到一个可能的结论（不是绝对确定，只是可能）：中子核的质量不可能小于1/10太阳质量。假如自然生成了比这还轻的中子核，它的引力还不够维持自己，压力将使它爆炸。

乍看起来，这并不排除太阳拥有一个中子核。毕竟，奥本海默和

塞伯估计允许的1/10个太阳质量的核太小了，可能会藏在太阳内部而不会对它的表面性质产生太多的影响（不会影响我们看见的事情）。但是，进一步计算核的引力与周围气体压力间的平衡，他们发现核的效应是藏不住的：核外包着一层白矮星型的物质壳，差不多跟太阳一样重，壳外却只有很少的正常气体，太阳将一点儿也不像我们现在看见的样子。所以，太阳不可能有中子核，维持太阳热量的能源一定来自别的地方。

那是哪儿呢？当奥本海默和塞伯在伯克利计算的时候，纽约绮色佳康奈尔大学的贝特（Hans Bethe）和乔治·华盛顿大学的克里奇菲尔德（Charles Critchfield）正在用新发现的核物理学定律来详细说明核燃烧（原子核聚变，卡片5.3）能够维持太阳和其他恒星的热量。爱丁顿对了，朗道错了——至少对太阳和大多数恒星来说是这样的。（90年代初，有些巨星看来也许就在用朗道的机制。[22]）

奥本海默和塞伯一点儿也不知道朗道的文章是他在绝望中写来躲避监狱和可能的死亡的，所以，在1938年9月1日，朗道正在布提斯卡雅监狱受折磨的时候，他们向《物理学评论》投了一篇批评他的文章。因为朗道是大物理学家，足以激起人们的热情，所以他们直截了当地说："[我们根据]朗道的估计……导出[中子核的]0.001个太阳质量的[极小]极限质量。这个数似乎是错的。……通常假定的自旋交换类型的[核力]否定了质量与太阳相当的恒星会存在[中子]核。"[23]

朗道的中子核与茨维基的中子星实际上是同一种东西。中子核不过正好是以某种方式处在正常恒星内部的中子星。在奥本海默看来，这一定是很清楚的，他现在开始考虑中子星了，当然也得无情面对茨维基本该解决却没能解决的问题：严格地说，大质量恒星在耗尽（据贝特和克里奇菲尔德的观点）维持其热量的核燃料后，会遭遇什么样的命运？它们会留下哪种遗骸呢：白矮星？中子星？黑洞？还是别的什么？

钱德拉塞卡的计算已经不容置疑地证明，质量小于1.4个太阳的恒星一定会变成白矮星。茨维基在大胆地猜想，至少某些质量大于1.4个太阳的恒星会坍缩形成中子星，并在此过程中产生超新星。茨维基能是正确的吗？是不是所有大质量恒星都像这样死亡，从而使宇宙远离黑洞？

奥本海默作为理论家的一大本领是能正确认识复杂的问题，能剥去它的复杂性，发现左右它的核心。几年后，他成为美国原子弹计划的领导人时，这种才能将卓越地发挥出来。现在，面对恒星的死亡，他感到忽略茨维基所宣扬的所有复杂的东西——如恒星坍缩的细节，正常物质向中子物质的转化，巨大能量的释放以及超新星和宇宙线的可能动力，这些都与恒星的最终命运无关。惟一相关的事情是中子星所能有的最大质量。假如中子星能有任意大的质量（上面图5.3中的曲线B），那么黑洞永远不会形成。假如中子星的质量有一个可能的极大值（图5.3曲线A），那么质量大于极大值的恒星可能在死亡时形成黑洞。

极大质量的问题已经完全清楚地提出来了，奥本海默还是跟平常的作风一样，和学生一道开始明确地、有条不紊地去解决它——这回跟他的年轻人叫沃尔科夫（George Volkoff）。奥本海默和沃尔科夫寻找中子星质量的经过和奥本海默在加州理工学院的朋友托尔曼的重要贡献，见卡片5.4。这个故事讲述了奥本海默的研究方法和物理学家采取的几个策略。那时，关于决定他们正在研究的现象的定律，他们知道一些，但不是全部：在这里，奥本海默知道量子力学和广义相对论，但不论他还是任何别的人都不太了解核力，尽管核力知识贫乏，奥本海默和沃尔科夫还是能够不容争辩地说明（卡片5.4），中子星有一个极大质量，它介于半个和几个太阳质量之间。

又经过50年的努力，到90年代，我们知道奥本海默和沃尔科夫是正确的；中子星的确有一个极大允许质量，现在知道它在1.5到3个太阳质量之间，[24]与他们的大致估计足一样的。另外，自1967年以来，

天文学家已发现了几百颗中子星，还高精度地测量了其中几个的质量，测得的质量都接近1.4个太阳质最，为什么呢？我们不知道。

卡片5.4　奥本海默、沃尔科夫和托尔曼的故事：寻找中子星质量[25]

在进行复杂性分析时，为帮助确定自己的方向，可以从较粗的“数量级”计算开始，这种计算只精确到一个因子，例如，10。奥本海默很熟悉这种经验方法，他先就是通过几页篇幅的粗略计算来看中子星是否能有一个极大质量。结果很诱人：他对任意中子星得到了6个太阳质量的极限。如果详细计算得到相同结果，奥本海默就可以下结论了：当恒星质量大于6个太阳时，黑洞就可能形成。

“详细计算”是说，为假想的中子星选一个质量，然后看在这个质量下恒星内的压力是否能与引力平衡。假如平衡可以达到，那么中子星就可以有这样的质量。这需要选一个个的质量，每次都寻求压力与引力间的平衡。这件事比看起来要艰难得多，因为压力与引力在星体内部必须处处平衡3不过，钱德拉塞卡在分析白矮星的时候就已经走过这条路了（分析是在爱丁顿的关照下，用爱丁顿的布伦瑞克计算器实现的，见第4章）。

奥本海默也能像钱德拉塞卡计算白矮星那样计算他的中子星，但必须先做两个关键的修改：第一，在白矮星，压力是电子产生的，而在中子星，压力由中子产生，所以物态方程（压力与密度的关系）将不相同；第二，在白矮星，引力较弱，用牛顿定律和爱因斯坦广义相对论都能很好地描述，两种描述会得到几乎完全一样的预言，所以钱德拉塞卡选择了简单的牛顿描述；反过来，在中子星，因为周长很小，引力太强，牛顿定律可能带来严重误差，所以，奥本海默只得用爱因斯坦的广义相对论定律来描述引力。1除了这两点变化——新的物态方程（用中子压力代替电子压力）和新的引力描述（用爱因斯坦的取代牛顿的）——奥本海默的计算与钱德拉塞卡的一样。

走到这里，奥本海默觉得可以把计算的细节交给学生了。他选择

了沃尔科夫，一个来自多伦多的年轻人，是1924年从俄国移民来的。

奥本海默向沃尔科夫解释了这个问题，告诉他，需要的引力的数学描述在托尔曼写的教科书《相对论、热力学和宇宙学》里。不过中子压力的状态方程更难一些，因为压力会受核力的影响（中子通过它相互推拉）。尽管在原子核内的密度下，我们对核力已有了很好的认识，但在大质量中子星深处中子可能面临的密度下，我们的认识还很贫乏。物理学家甚至不知道，在这种密度下核力是吸引的还是排斥的（中子是相互推还是拉），所以也就没有办法知道，核力是增大压力还是减小压力。但奥本海默有对付这些未知事情的办法。

奥本海默建议沃尔科夫先假定核力不存在，那么压力将是我们很好地认识了的那一类，就是中子简压力（由中子的"幽闭"运动产生的力）。平衡中子简并压力与引力，然后根据平衡计算中子星在没有任何核力的宇宙中可能具有的结构和质量。然后，在核力以这样那样方式作用的真实宇宙中，估计星体的结构相质量会如何变化。

在这么好的指引下，不会再有迷失了。沃尔科夫每天同奥本海默讨论，加上托尔曼教科书的帮助，他只用几天就导出了中子星内部引力的广义相对论描述，也只用几天就将大家熟知的简并电子压力的状态方程转化成了简并中子压力的状态方程。平衡压力与引力，沃尔科夫得到一个复杂的微分方程，它的解将告诉他星体的内部结构。这时，他被困住了。沃尔科夫竭尽全力也解不了他的微分方程，得不到星体结构的公式；他只好像钱德拉塞卡计算白矮星那样数值求解他的方程。1934年，钱德拉塞卡曾在爱丁顿的布伦瑞克计算器上敲了好多天的键钮来计算类似的白矮星结构，同样，沃尔科夫在1938年11月到12月间的大部分时间也费在马尔琴特（Marchant）计算器的键钮上了。

沃尔科夫在伯克利埋头计算时，帕萨迪纳的托尔曼正在采取不同的策略；他更喜欢用公式而不是计算器的数字来表达恒星的结构。一

个公式可以表现包含在许许多多数表里的信息。如果他得到了正确公式，它会同时包含1个太阳、2个太阳、5个太阳质量——以至任意质量的恒星的结构。但托尔曼凭他卓越的数学技巧，也没能用公式解决沃尔科夫的方程。

“另一方面”，托尔曼大概会跟自己争论，“我们知道沃尔科夫用的并不真的是那个正确的物态方程。沃尔科夫忽略了核力；因为我们不知道那个力在高密度下的情况，所以也就不知道正确的物态方程。那么让我来问一个与沃尔科夫不同的问题：中子星的质量如何依赖于物态方程呢？假定物态方程很‘硬’，就是说，它给出特别高的压力，我要问，在这种情况下中子星的质量有多大？接下来，我假定物态方程很‘软’，就是说，它给出特别低的压力，那么这个时候中子星质量又如何呢？在每种情况下，我都将假想的物态方程调整到我能用公式解沃尔科夫微分方程的形式。尽管我用的物态方程几乎肯定不会是正确的，但我的计算还是能够一般性地告诉我，当自然选择了硬物态方程时，中子星质量可能是多大；当自然选择了软方程时，它又可能是多大。”

10月19日，托尔曼给奥本海默发了封长信，讲了他根据几个假想物态方程导出的星体结构公式和中子星质量。大约一个星期后，奥本海默驱车来到帕萨迪纳，同托尔曼就这个问题讨论了几天。11月9日，托尔曼又给奥本海默写了封长信，公式更多了。[26]同时，沃尔科夫还在他的马尔琴特键钮上敲打。12月初，他算完了。他得到了质量为0.3，0.6，0.7个太阳质量的中子星的数值模型，发现如果我们的宇宙中没有核力，那么中子星将总是低于0.7个太阳质量。

太奇怪了！奥本海默在沃尔科夫计算前的粗略估计是6个太阳质量。为了阻止大质量恒星成为黑洞，详细计算本应将极大质量推到100个或更多的太阳质量，相反，它却把质量降下来了——只有0.7个太阳质量。

托尔曼来到伯克利了解详情。50年后，沃尔科夫高兴地回忆了当时的情景："我记得自己很怕向奥本海默和托尔曼解释我做的事情。我们坐在伯克利旧教工俱乐部的草地上。在高高大树下的茵茵绿草间是令人尊敬的两位先生，还有我这位刚读完博士的研究生，要向他们解释我的计算。"[27]

既然已经知道了没有核力的理想宇宙的中子星的质量，奥本海默和沃尔科夫现在可以估计核力的影响了。托尔曼为不同假想物态方程仔细求解的公式将在这儿发挥作用。从托尔曼的公式可以大概看到，如果核力是排斥的从而物态方程比沃尔科夫用过的"硬"，星体结构将如何改变；如果核力是吸引的从而物态方程更"软"，它又将如何。在可信的核力范围内，这些变化都不大。托尔曼、奥本海默和沃尔科夫最后认为，中子星还是必须有一个极大质量，处在大约半个到几个太阳质量之间。[28]

奥本海默和沃尔科夫的结论不会令爱丁顿和爱因斯坦那些诅咒黑洞的人满意。如果人们最终相信了钱德拉塞卡（在1938年，大多数天文学家都逐步理解他了），相信了奥本海默和沃尔科夫（那时反驳他们也不容易），那么，不论白矮星还是中子星，它们那样的墓穴都不能埋葬大质量的恒星。还有别的可信的方式让大质量恒星避免死于黑洞吗？是的，有两条。

第一，所有大质量恒星都可能在成长中释放大量的物质（例如，通过强大的风吹散星球表面，或者通过核爆炸），从而将质量减小到1.4个太阳以下，进入白矮星的墓穴；或者，（假如谁相信茨维基的机制，不过很少有人相信）它们会在超新星爆发中释放物质，将质量减到大约1个太阳质量以下，然后终结在中子星墓穴里。从40年代到50年代直到60年代初，大多数天文学家——假如他们都考虑这个问题——都相信这一点。

第二，除白矮星、中子星和黑洞墓穴外，可能还存在着大质量恒

星的第四种墓穴，是30年代还没认识到的。例如，我们可以想象图5.3的一个墓穴，周长在中子星和白矮星之间——几百或1 000公里。大质量恒星可能在变得更小而形成中子星或黑洞之前就在这种墓穴停止了收缩。

假如没行第二次世界大战和后来冷战的干扰，奥木海默和他的学生们或别的人，很可能在40年代考察这种墓穴，他们将严格地证明，不存在这第四种墓穴。

然而，二次大战确实爆发，它耗尽了几乎全世界所有理论物理学家的精力，接着，战后的氢弹紧急计划进一步耽误了物理学家回到正常生活中来（见下一章）。

最后，50年代中期，两个物理学家从各自的氢弹研究中脱颖而出，来到奥木海默和他的学生们离开了的地方。他们是美国普林斯顿大学的惠勒（John Archibald Wheeler）和莫斯科应用数学研究所的泽尔多维奇（Yakov Borisovich Zcl'dovich）——两位杰出的物理学家，将成为本书后面的主角。

[1] 关于不同的物理学描述和它们的有效范围的关系，请看第1章最后一节的讨论（“物理学定律的本质”）。

## 惠勒

1956年3月，惠勒花了几天时间研究钱德拉塞卡、朗道以及奥本海默和沃尔科夫的文章，他在这儿发现了值得深入探索的奥秘。[29]质量大于1.4个太阳的恒星死亡时只能形成黑洞，而没有别的选择，这能是真的吗？不久后，惠勒写道，“在广义相对论关于宇宙的结构和演化的所有结果中，大质量物体的命运问题是最具挑战性的。”他决心去完成钱德拉塞卡、奥本海默和沃尔科夫开创的星体墓穴的探索。

惠勒为了把他的使命表达得更准确，仔细描述了构成冷星和死星的那一类物质，称它为热核演化终点的物质，因为热核一词在为星体的核燃烧和氢弹提供动力的聚合反应中已经用得很普遍了。这类物质是绝对冷的，已经燃尽了核燃料，不会有什么办法通过任何类型的核反应从它的核内得到更多的能。因为这个理由，本书将用冷死物质来代替“热核演化终点的物质”。



J·A·惠勒，约1954年[Blackstone-Shelburne摄于纽约；惠勒提供。]

惠勒为自己设的目标是认识所有能用冷死物质构造的物体，这将包括像铁球那样的小物体，像由铁构成的冷死行星那样的较重物体以及其他更重的物体，如白矮星、中子星和别的物理学定律允许的任意类型的冷死物体。惠勒想要一个冷死事物的综合编目。

惠勒的工作模式同奥本海默一样，身边围着一帮学生和博士后。他让他们中的一个来自犹他州的虔诚摩门教徒哈里森（B.Kent Harrison）来解决具体的冷死物质状态方程。这个物态方程将描述这类物质的压力在密度越来越高时会如何增加——或者等价地说，它的

压缩阻抗如何随密度的增长而变化。

惠勒为哈里森计算冷死物质的状态方程准备了足够的指导，因为在决定物质结构的物理学定律（量子力学定律和核物理定律）的领域内，他跻身于世界最伟大的专家行列中。在过去的20年里，他发展了描述原子核行为的强有力的数学模型；他同玻尔一道发现了核裂变（像铀、钚那样的重原子的分裂，这是原子弹的基础）；他还曾是设计美国氢弹的一个小组的领导人（第6章）。[30]凭着这些经历，惠勒指引哈里森在错综复杂的问题中穿行。

他们关于冷死物质状态方程的分析结果，我们在卡片5.5中讨论说明。在白矮星密度，它与钱德拉塞卡研究白矮星时用的物态方程是一样的（第4章）；在中子星密度，它与奥本海默和沃尔科夫用过的相同（卡片5.4）；密度在白矮星以下以及在白矮星和中子星之间时，方程是全新的。

掌握了冷死物质的状态方程后，惠勒请来自日本的博士后若野正巳（Masami Wakano）去做一件沃尔科夫为中子星和钱德拉塞卡为白矮星做过的事情：将物态方程与广义相对论方程结合起来，描述星体内部引力与压力的平衡；根据这样的结合导出描述星体结构的微分方程；然后数值求解这个微分方程。数值计算将给出所有冷死星体的内部结构情况，而最重要的是能给出星体的质量。

30年代，为了计算一个星体的结构（星体内部密度、压力和引力的分布），钱德拉塞卡和沃尔科夫在剑桥和伯克利的计算器上敲打了好多天。50年代就大不一样了。普林斯顿有世界上第一台数字计算机，MANIAC——在普林斯顿高等研究院的一间满是真空管和电缆的屋子里，这原是为氢弹设计建造的。有了MANIAC，若野用不了一个小时就能解决一颗星的结构。

卡片5.5　冷死物质状态的哈里森-惠勒方程[31]

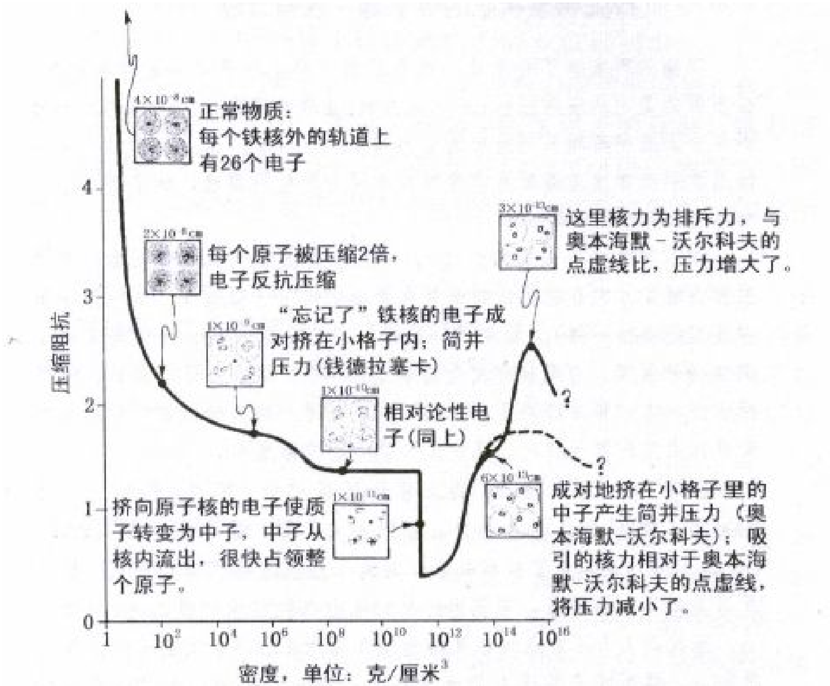
下面的图表示了哈里森-惠勒物态方程。水平画的是物质密度，竖直画的是它的压缩阻抗（或绝热指数，物理学家喜欢这么叫它）——1个百分点的密度增长所对应的压力增长的百分数。曲线上的格子说明物质在从低密度压缩到高密度时在微观上发生的事情。格子的大小写在顶部（以厘米为单位）。

在正常密度下（图的左边），冷死物质由铁构成。如果物质的原子核比铁重，它会通过分裂释放能量而变成铁（核裂变，与在原子弹中发生的事情一样）。如果核比铁轻，它们会通过结合释放能量而变成铁（核聚变，与氢弹中发生的事情一样）。物质一旦形成铁，就不再以任何方式释放核能了。物质形成铁核时，核力会比在它们形成其他任何类型的原子核时，将中子和质子束缚得更紧。

在铁从7.6克/厘米$^3$的正常密度压缩到100克/厘米$^3$、再到1000克/厘米$^3$时，它也像岩石那样反抗压缩：每个电子以“幽闭”（简并式的）运动来抵抗相邻原子的电子挤压。阻抗起初很大，但不是因为排斥力特别强，而是因为初始压力在低密度时很低。（回想一下，阻抗是1个百分点的密度增长所伴随的压力增长的百分点。压力很低时，猛烈增长的压力代表着巨大的增长百分点。也就是巨大的“阻抗”。过后，在压力大的高密度下，强烈的压力增长代表着小得多的增长百分点，从而阻抗也小得多。）

起初，冷物质被压缩，电子紧密聚集在铁核周围，形成电子轨道构成的电子云。（每个轨道上实际有两个电子，而不是一个——第4章忽略了这点微妙的不同，但在卡片5.1中简单讨论过了。）随着压缩继续，每个轨道和它的两个电子被逐渐限制在一个越来越小的活动空间，为了反抗这种限制，幽闭的电子变得更像波一样不规则地高速运动（“简并运动”，见第4章）。当密度达到$10^5$克/厘米$^3$时，电子的简并运动和它产生的简并压力会变得非常大，完全超过了原子核作用在电子上的电力。电子不再围绕铁核，根本不把它放在眼里。原先还是一块铁的冷死物质，现在成了白矮星的构成材料，物态方程也变成

了钱德拉塞卡、安德森和斯托纳在30年代初计算过的那一个（图4.3）：5/3的阻抗光滑地变到4/3，这时候，电子不规则运动的速度接近光速，密度为$10^7$克/厘米$^3$。




根据哈里森和惠勒的计算，白矮星物质向中子星物质的转变发生在$4\times10^{11}$克/厘米$^3$的密度计算表明了转变的几个阶段：第一阶段，电子被挤入原子核，核的质子吞没电子而形成中子。物质从而失去了一些维持压力的电子，压力阻抗突然变小，这导致了物态方程曲线的陡落（见上图）。随着第一阶段的进行，阻抗陡然下降，原子核因中子而越胀越大，也触发了第二阶段：中子开始从核中流出（被挤出来），汇入留在核外的少数电子。流出的中子跟电子一样，凭自己的简并压力对抗着不断的挤压。中子简并压力终止了物态方程的陡落，压力阻抗又重新开始上升。在第三阶段，密度为$10^{12}\sim4\times10^{12}$克/厘米$^3$，每个因中子而膨胀的核都彻底破裂了，碎成一个个中子，形成

奥本海默和沃尔科夫研究过的中子气，以及少数散开的电子和质子。密度从此开始上升，物态方程表现出奥本海默-沃尔科夫中子星的形式（忽略核力时，即图中的虚线；实线考虑了90年代对核力影响的最新认识）。

若野的计算结果见图5.5。本图是冷死物体的严格而最终的分类，它回答了我们在本章前面图5.3的讨论中所提出的所有问题。

在图5.5中，星体周长向右，质量向上。周长和质量处在白区的任何恒星的内部引力都大于其压力，所以引力使星体向左收缩。在阴影区的恒星，压力大于引力，所以压力使星体向右膨胀。只有在白区和阴影区的界线上，引力和压力才相互平衡，因此，这条边界线就是处于压力-引力平衡状态的冷死星曲线。

假如你沿平衡曲线追踪，你会遇到密度越来越高的死“星”。在最低密度（沿图的底线，大部分看不见），这些“星”根本不是什么恒星，不过是由铁构成的冷行星。（当木星最终耗尽内部的热辐射而冷却下来时，也将落在平衡曲线最右端的附近，尽管它大部分是由氢而不是铁构成的。）比行星密度高的地方是钱德拉塞卡的白矮星。

当你到达曲线上白矮星部分的最高点（具有1.4个太阳的钱德拉塞卡极限质量的白矮星[1]），然后继续向更高密度走，你会遇到不可能自然存在的冷死星，因为它们在收缩和膨胀中是不稳定的（卡片5.6）。从白矮星密度走到中子星密度，这些不稳定平衡星体的质量将减小到大约0.1个太阳质量的极小值，周长为1000千米，中心密度为$3\times10^{13}$克/厘米$^3$。这是最初的中子星，也就是奥本海默和塞伯研究过的“中子核”，他们曾证明这种核不可能比0.001个太阳质量更小，那是朗道为太阳内核所设想的质量。

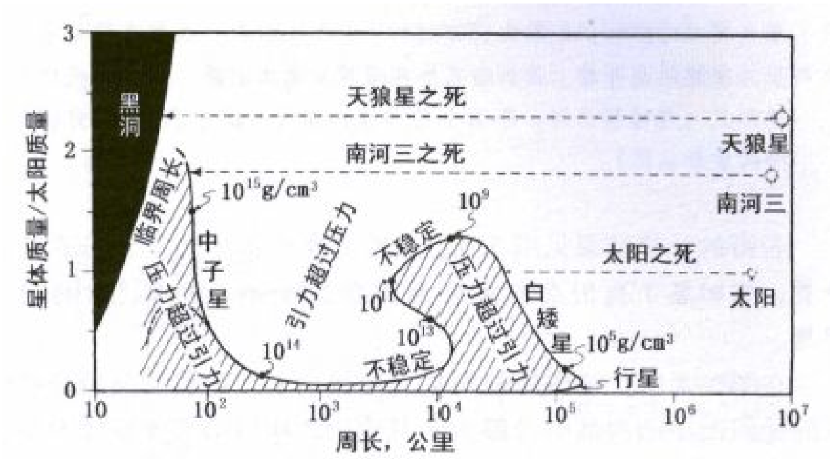



图5.5　冷死星的周长（水平方向）、质量（垂直方向）和中心密度（标在曲线上）。这是若野在惠勒指导下用卡片5.5中的物态方程计算的结果。中心密度超过原子核（$2\times10^{14}$克/厘米$^3$）时，实线是90年代的结果，恰当考虑了核力的影响；虚线是奥本海默和沃尔科夫忽略核力的结果。[32]

## 卡片5.6　白矮星和中子星之间的不稳定过客

图5.5的平衡曲线上，所有处在白矮星和中子星之间的星体都是不稳定的。例如，中心密度为$10^{13}$克/厘米$^3$的星体（它的质量和周长在图5.5中标记$10^{13}$的点上）。在$10^{13}$这一点，星体是平衡的，它的引力和压力彼此完全平衡。然而，它像立在尖儿上的铅笔一样是不稳定的。

假如某个随机力（例如星际气体落在星体上）将星体轻轻挤了一下，就是说减小了它的周长，使它在图5.5中向左移了一点，进入白区，那么，星体的引力将开始超过压力，把星体引向坍缩；星体坍缩时，会强烈地向左移动，穿过中子星曲线进入阴影区，而中子压力将在这里暴涨，阻止坍缩，把星体表面向外推，使它落回中子星曲线，进入中子星的墓穴。

反过来，假如在$10^{13}$那点的星体不是被小的随机力向内挤压，而是向外推了一点（例如，某些中子的不规则运动随机增大了），那么

它将进入压力超过引力的阴影区；压力会使星体表面爆炸，向外穿过白矮星曲线进入图的白区，在那儿，引力将占上风，把它向内拉回白矮星曲线，进入白矮星的墓穴。

这种不稳定性（$10^{13}$的星体，收缩一点会坍缩成为中子星；扩张一点会爆炸成白矮星）意味着，在$10^{13}$的密度上——或者在平衡曲线上标明“不稳定”的区间内，不可能长时间存在什么真实的星体。

走在平衡曲线上，我们遇见了质量大约从o.1～2个太阳质量的整个中子星族。2个太阳的中子星极大质量到90年代仍然有些不确定，因为极高密度下的核力行为还没有得到很好的认识。极大值可能低到1.5个太阳，但不会更低；也可能高到3个太阳，但也不会更高多少。

在平衡曲线的（近似）2个太阳质量的峰值上，中子星终结了。当我们沿着曲线进一步向更高的密度追踪时，平衡的星体也像在白矮星和中子星之间那样，是不稳定的（卡片5.6）。因为相同的理由，这些不稳定的“星”在自然界是不存在的。假如真形成了这种星体，它们立刻会坍缩成为黑洞，或爆炸成为中子星。

图5.5是绝对严格和不容争辩的：在白矮星和中子星之间不存在第三类稳定的大质量的冷死物体。因此，像天狼星那样质量大于2个太阳的恒星在耗尽核燃料后，要么释放所有多余的质量，要么发生坍缩，超过白矮星和中子星的密度，进入临界周长以内——在90年代的今天我们能完全肯定，它们会形成黑洞。坍缩是必然的。对质量足够大的星体来说，不论电子的简并压力还是中子间的核力，都阻止不了坍缩。引力甚至超过了核力。

不过，还有一条出路，可以让所有恒星，甚至最大质量的恒星，逃脱黑洞的命运；也许，所有大质量恒星在它们晚年或死亡中（以风或爆炸的形式）放出足够的质量，使自己小于2个太阳质量，这样就能终结在中子星或白矮星的墓穴。在40年代、50年代和60年代初，

假如天文学家从各方面考虑星体最终命运的问题，他们都会倾向这个观点。（然而，大体说来，他们不会考虑这个问题。没有什么观测数据促使他们去考虑它，天文学家的注意力都被吸引到其他类型的物体上去了——正常恒星、星云、星系——他们得到的观测数据已经够多了，够富挑战性了，够报答他们了。）

在90年代的今天，我们知道重星确实会在成长和死亡时放出大量物质。事实上，放出物质之多是惊人的：大多数诞生时有8个太阳质量的恒星因释放了足够的质量而终结在白矮星的墓穴；而大多数天生有8~20个太阳质量的恒星，最后成了中子星。这样看来，大自然也几乎不愿自己看到黑洞。但并不完全这样：大多数观测数据提醒我们（但没有确实证明），多数天生大于20个太阳质量的恒星在死亡时还是那么重，它们的压力抗拒不了引力。当它们耗尽核燃料开始冷却时，引力超过了压力，它们便坍缩而形成黑洞。在第8章里，我们会看到这样一些观测数据。

我们还有好些关于科学和科学家的本质的东西，需要从30年代的中子星和中子核的研究中学习。

奥本海默和沃尔科夫研究的东西是茨维基的中子星，而不是朗道的中子核，因为中子星没有星体物质的包围。不过，奥本海默对茨维基不太尊重，他不愿用茨维基的名字来命名它们，而坚持用朗道的名字。这样，他和沃尔科夫讲述他们结果的那篇发表在1939年2月15日《物理学评论》上的文章，题目就叫“关于大质量中子核”。[33]为保证没人误会他关于这些星体思想的来源，奥本海默在文章里零星地提了朗道，而茨维基先发表的那么多东叫，他一次也没引用过。

就茨维基来说，1938年，他一直在关注着托尔曼、奥本海默和沃尔科夫的中子星研究，他们怎么做这个？他愤怒了。中子星是他的孩子，而不是他们的；他们没有权力研究中子星——而且，尽管托尔曼偶尔跟他谈谈，奥本海默却根本没理他！

然而，在茨维基就中子星写的大量文章中，只有空谈和猜想，没有实质性的内容。他更多地在忙着观测寻找超新星（很成功），忙着写文章谈中子星和它在超新星中的作用，从来没找时间来充实内容。不过，想同别人争，还得自己行动。1938年初，他尽自己的努力完成了中子星的数学理论，并同他的超新星观测结合起来。他努力的结果发表在1939年4月15日的《物理学评论》上，题目是“高坍缩星体的观测和理论”。[34]他的文章比奥本海默和沃尔科夫的长两倍半，没有单独提他们两个目前的文章，不过提了沃尔科夫个人的一篇辅助性的小文章。这篇文章没有什么值得记住的东西，事实上，很多都是完全错误的。相反，奥本海默-沃尔科夫一文却是一篇杰作，优美而富有远见，所有细节也都是正确的。

尽管如此，半个多世纪过去了，我们今天还是尊敬茨维基的——他发现了中子星的概念；正确认识到了中子星是超新星爆发的产物和能源；他和巴德在观测上证明了超新星实际上是一类独特的天体；他开创并实现了几十年的确定性的超新星观测研究；另外，他还有许多中子星和超新星之外的见解。

他对物理学定律的认识那么少，而他的远见那么多，这是怎么回事呢？在我看来，他的身上融合着几种个性特征：他对理论物理学的理解足够让他在定性上（如果不是定量的话）正确认识事物；强烈的好奇心令他紧跟发生在物理学和天文学中的每一件事情；他能以某种直觉的方法识别（别人很少能做到）不同现象之间的联系；另外，同样重要的是，他太相信内心的通向真理的道路，从来不怕他的猜想会如何愚弄自己。他知道他是对的（尽管他常常是错的），如山的证据也不能令他相信他认为是错误的东西。

朗道跟茨维基一样，也很自信，也不怕自己像傻子。例如，他毫不犹豫地发表了他1931年的观点：恒星由超致密的星核提供能源，量子力学定律在那儿失败了。朗道对理论物理学的把握远远超过了茨维基，他是20世纪十大理论家之一。不过，他的猜想错了，而茨维基是

对的。太阳没有中子核的能源；超新星才由中子星提供能量。那么，朗道与茨维基相比，是不是差点儿运气呢？也许部分是这样，但还有一个因素：茨维基投身在威尔逊山的环境中，那儿是世界天文观测的中心。他还与世界大实测天文学家巴德（他掌握着观测数据）合作；在加州理工学院，他可以而且确实几乎每天都与世界上的宇宙线观测大师们交谈。反过来，朗道与实测天文学几乎没有直接往来，他的文章也证明了这一点。没有这些实际接触，他不可能培养出对远在地球之外的那些事物的敏锐感觉。朗道最伟大的胜利是用量子力学定律解释超流现象，在这项研究中，他与实验家卡皮查进行过广泛的交流，那时卡皮查正在探索超流的细节。

与茨维基和朗道不同的是，对爱因斯坦来说，观测与理论之间的密切接触并不很重要：他发现广义相对论的引力定律就几乎没有靠什么观测。但这是一个罕见的例外。观测与理论在多方面的相互影响对物理学和天文学大多数分支的发展是有基本意义的。

奥本海默又如何呢？他的物理学造诣是堪与朗道比肩的。他与沃尔科夫合作的关于中子星结构的文章是历史上最伟大的天体物理学文献之一，但尽管文章优美，却“只不过”为中子星的概念填充了一些细节。概念实际上还是茨维基的孩子——超新星也是他的，星核坍缩形成中子星从而为超新星提供动力的思想还是他的。为什么奥本海默有那么多好条件，却没有茨维基那么多创造呢？我想，主要是因为他不愿意——甚至也许害怕——猜想。奥本海默的好朋友和崇拜者拉比（Isidore I.Rabi）更深刻地描述过这一点：

“我以为，奥本海默在某些方面受科学传统以外的东西的影响太深，比如他对宗教特别是对印度宗教的兴趣，产生了雾一般的对宇宙奥秘的感觉。面对已经做过的事情，他把物理学看得很清楚，但在学科的边缘，他却感到神秘和奇异的事情比实际存在的多得多。他不太相信已经掌握的理论工具的威力，没有将他的思想发挥到尽头，因为他本能地感到，如果他和他的学生想比现在走得更远，就必须靠新的

思想和方法。”[35]

[1] 实际上，图5.5（若野的计算）的白矮星极大质量是1.2个太阳，比钱德拉塞卡计算的值略小。区别在于不同的化学组成：若野的星体由“冷死物质”（大部分是铁）构成，电子数是核子（中子和质子）数的46%，钱德拉塞卡的星由氦、碳、氮、氧等元素构成，电子数占核子数的50%。事实上，我们宇宙间大多数白矮星更像钱德拉塞卡的，不太像若野的。这也是为什么我在本书一直采用钱德拉塞卡的极限质量值：1.4个太阳。

# 第6章　坍缩成什么？

所有的理论物理学家

都挡不住一个结论：

坍缩产生黑洞

不是冤家不聚头。奥本海默和惠勒，这两个在宇宙和人类处境问题上观念截然不同的人，越来越发现他们在国家安全和核武器政策等深层问题上也站在两个极端——而他们现在的问题是黑洞。

他们相聚在比利时布鲁塞尔大学的演讲厅。这两位在新泽西普林斯顿的邻居，与来自世界各地的其他31位大物理学家和天文学家一起，到这儿来聚会一个星期，讨论宇宙的结构和演化。

那是1958年6月10日，星期二。[1]惠勒刚向聚在这儿的著名学者们报告了他与哈里森和若野最近的计算结果——这些计算不容争辩地确认了所有可能冷死星体的质量和周长（第5章）。他填补了钱德拉塞卡和奥本海默-沃尔科夫计算之间的鸿沟，还证实了他们的结果：质量大于2个太阳的恒星在死亡时，坍缩是必然的，而且，只要死亡中的恒星没有释放出足够的质量将自己减轻到大约2个太阳的极限质量以下，它的讲缩就不会产生白矮星、中子星或者任何其他类型的冷死星体。

“在广义相对论关于宇宙的结构和演化的所有结果中，大质量物体的命运问题是最具挑战性的。”惠勒说。在这一点上，听众们都会同意。接着，惠勒似乎回答了24年前爱丁顿对钱德拉塞卡的攻击（第

4章）。他描述了奥本海默关于大质量恒星一定会通过坍缩成黑洞而死亡的观点，然后反驳说，这类坍缩“不会得到令人接受的答案”。为什么呢？基本上还是当年爱丁顿拒绝它的那个原因，用爱丁顿的话说，“应该存在一个自然律来阻止恒星那么荒谬的行为。”但是爱丁顿与惠勒之间存在着深刻的差别：爱丁顿1935年猜想的让宇宙摆脱黑洞的机制立刻就被玻尔那样的专家否定了；惠勒1958年猜想的机制一时还证实不了，也否定不了——15年后发现它部分是正确的（第12章）。

惠勒的猜测是这样的。（在他看来）形成黑洞的坍缩在物理学上一定是不合理的，应该抛弃，因此“似乎没有什么能摆脱这样的结论：坍缩恒星中心的核子[中子和质子]必然会融人辐射，而辐射会很快逃离恒星以减少恒星的质量[低于2个太阳]”从而使它能终结在中子星的墓穴。[2]惠勒愿意承认，这样的核子向辐射的转化超出了已知物理学定律的范围。不过，这种转化可能是那时还不知道的广义相对论与量子力学“结合”的一个结果（12～14章）。对惠勒来说，这才是“大质量问题”最诱人的一点：既然坍缩成黑洞是荒谬的，他只好考虑一种全新的物理学过程（图6.1）。

奥本海默没有被感动。惠勒讲完后，他头一个站起来，带着年轻时少有的礼貌，强调了自己的观点：“我不知道远比太阳重的非旋转物质是否真的能在星体演化过程中出现；但如果有的话我相信它的坍缩可以在广义相对论框架内进行描述[而不需要什么新的物理定律]。这些物质持续经受引力收缩，最终越来越多地与宇宙的其余部分隔绝[也就是形成黑洞]，这难道不是最简单的设想吗？”（见图6.1。）

奥本海默-斯尼德观点：

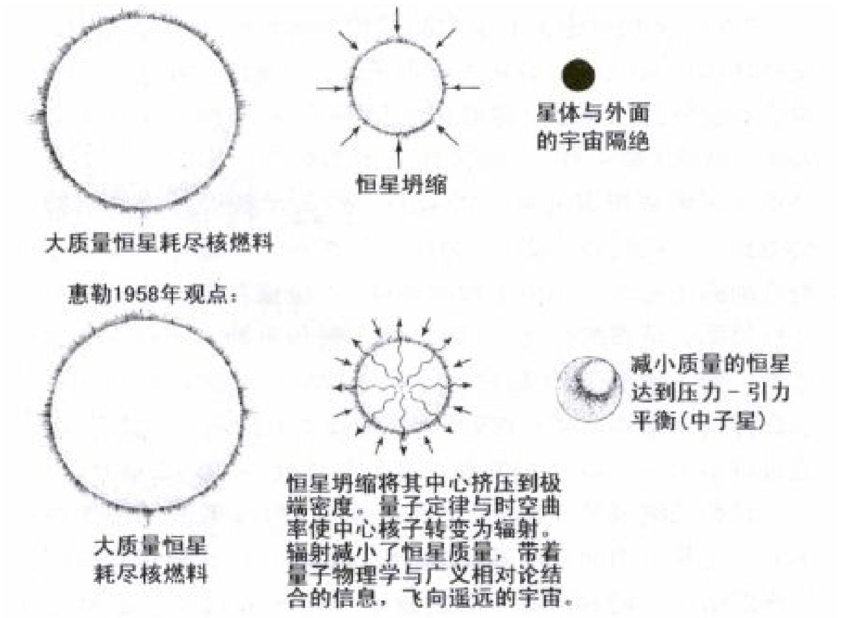



图6.1 关于大质置恒星命运的奥本海默观点（上序列）与惠勒1958年观点（下序列）的比较。

惠勒也一样有礼貌，但还是坚持着自己的立场。“很难相信‘引力隔绝’是令人满意的答案”，他这么说。

奥本海默对黑洞的信服，来自他19年前的具体计算。

## 黑洞的诞生：初步认识

1938年冬到1939年，奥本海默刚与沃尔科夫完成中子星的质量和周长的计算（第5章）。他确信，大质量恒星在死亡时会发生坍缩。接下来的挑战是显然的：用物理学定律来计算坍缩的细节。从围绕恒星的轨道上看，坍缩像什么？从恒星表面看，坍缩像什么？坍缩千百年后，恒星的最终状态是什么？

计算是不会容易的。奥本海默和他的学生们面临着巨大的数学挑战：坍缩的恒星会随时间改变性质，而奥本海默-沃尔科夫中子星却是

静止不变的。在坍缩恒星的内部，时空曲率会变得很大，而在中子星内它却很小。为处理这些复杂的事情，还需要一个很专门的学生，人选是明摆着的：哈特兰·斯尼德（Hartland Snyder）。

斯尼德跟奥本海默的其他学生不同。别人来自中产阶级家庭，而斯尼德是工人。在伯克利有人说他成为物理学家前曾是犹他州的一个卡车司机。据塞伯回忆，“奥比的学生习以为常的好多事情，如听巴赫和莫扎特，玩儿弦乐四重奏，吃精美的食品，谈自由政治，哈特兰都瞧不起。”[3]

加州理工学院的那群核物理学家比奥本海默周围的人热闹得多；每年春天奥本海默来帕萨迪纳时，哈特兰正好也到这儿来。福勒说，“奥比修养极高，爱文学、艺术和音乐，还懂梵文，而哈特兰则跟我们这群闲人差不多。他喜欢克洛格实验室的晚会。晚会上，托米（Tommy Lauritsen）弹钢琴，查理（Charlie Laurit-sen）[实验室的头儿]拉小提琴，我们大家喝着酒，唱着校歌和酒歌。在奥比所有的学生中，哈特兰是最特别的一个。”[4]

在智力上斯尼德也与众不同。塞伯回忆，“哈特兰在数学难题上比我们大家都能干，我们做的粗略计算他都能有很好的进。”[5]正因为这些才干，他成为当然的做坍缩计算的人。

在展开完全的复杂计算前，奥本海默坚持（像往常一样）先对问题作一个大概的考察。[6]只花一点儿气力，能认识多少？初步考察的关键是恒星外的弯曲时空的史瓦西几何（第3章）。

史瓦西通过解爱因斯坦广义相对论场方程发现了他的时空几何。这是既不坍缩，也不爆炸，也不脉动的静态恒星的外部解，但是，哈佛大学数学家贝克霍夫（George Birkhoff）在1923年证明了一个重要的数学定理：史瓦西几何描述了任何球状恒星的外部，不仅包括静态的恒星，也包括坍缩的、爆炸的和脉动的恒星。

初步计算时，奥本海默和斯尼德简单假定球状恒星在耗尽核燃料后会无限坍缩，他们没考虑恒星内部会发生什么事情，而去计算在远处的人看来，坍缩的恒星会像什么样子。他们很轻松地发现，由于坍缩恒星外的时空几何与任何静态恒星外的几何是一样的，所以坍缩的恒星看起来很像一个静态恒星序列，序列中每颗星都比前一颗更紧密。

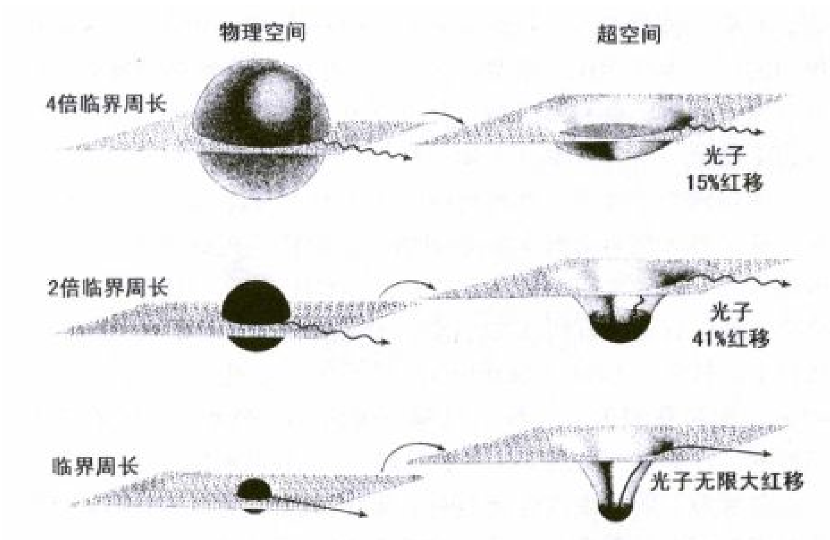



图6.2　（与图3.4相同）广义相对论对空间曲率和光线红移的预言。三个高度致密的静态（不坍缩的）恒星，有相同的质量和不同的周长。

到1920年前后，这种静态恒星的外在表现已经研究过20多年了。图6.2画出了我们在第3章曾用来讨论那些表现的嵌入图。回想一下，每个嵌人图都刻画了恒星内部和附近的空间曲率。为了综合表现这些曲率，图中只画出了三维空间的二维曲率，也就是落在恒星赤道“面”的那两维（图左）。想象将赤道面从恒星拉出来，从我们和它所在的物理空间拉出来，将它放到一个平直的（没有弯曲的）假想超曲面中，我们就能具体看见空间的曲率了。在没有弯曲的超曲面中，赤道面的弯曲几何只有通过向下弯曲的碗来表现（右图）。

图中的3个静态恒星构成的序列模拟了奥本海默和斯尼德准备分析的坍缩过程。每颗星的质量相同，但周长不同。第一颗星约是临界周长的4倍（也就是恒星在引力作用下形成黑洞的周长的4倍）。第二颗星是2倍。第三颗星正好具有临界周长。嵌入图表明，恒星越接近临界周长，周围空间的曲率就越大。然而，曲率不会变得无限大。即使恒星处在临界周长，碗状几何也是处处光滑的，没有尖点，也没有褶皱。也就是说，时空曲率不是无限大的；于是相应地，潮汐引力（将人从头到脚拉长或引起地球潮汐的那种力）作为时空曲率的物理学表现，在临界周长也不会无限大。

我们在第3章还讨论了从静态恒星表面发出的光的命运。我们知道，因为星体表面的时间比远处的时间流得更慢（引力的时间膨胀），所以在远处接收的来自恒星表面的光将具有被拉长了的振动周期，相应地也就有更大的波长和更红的颜色。光从恒星的强大引力场中往外“爬”时，它的波长将向光谱的红端移动（引力红移）。当静态恒星比临界周长大4倍时，光的波长被拉长了15%（见图右上角的光子）；当恒星为2倍临界周长时，红移为41%（中间）；当恒星刚好处在临界周长时，光的波长将产生无限红移，意味着光根本没有留下能量，从而也就不存在了。

奥本海默和斯尼德在大概的计算中根据这个静态恒星序列发现了两件事情：第一，像这些静态恒星一样，坍缩的恒星在接近临界周长时可能会产生强大的时空曲率，但不会无限大，因而也不会有无限大的潮汐引力。第二，恒星坍缩时，表面发出的光将经历越来越大的红移，当恒星达到临界周长时，红移将无限大，恒星也就完全看不见了。用奥本海默的话来说，我们将看着恒星将它自己与外面的宇宙“隔绝”开来。

奥本海默和斯尼德问自己，恒星的内部性质——在大概计算时忽略了——能有什么办法使恒星摆脱隔绝的命运吗？例如，坍缩会不会因某种力量被迫慢下来，即使经过无限长时间也不可能实际达到临界

周长？

他想通过真实恒星坍缩（如图6.3左图所示）的具体计算来回答这些问题。任何真实恒星都像地球一样会自转，至少也会慢慢地转，自转的离心力会使恒星赤道像地球赤道一样，至少会向外凸出一点儿。所以，恒星不可能是正球形的。恒星在坍缩时会像溜冰者收回双臂那样越转越快，这使星体内部的离心力越来越大，赤道隆起将越来越显著——也许会显著到令坍缩停止的地步，这时向外的离心力与引力完全平衡了。任何真实恒星在中心都有很高的密度和压力，而外层的密度和压力都较低；当它坍缩时，高密度会像浆果饼里的蓝色浆果一样到处成堆。而且，恒星的气体物质在坍缩时会产生激波——像破碎的海浪——从恒星表面的某些部分射出物质和质量，就像海浪溅起浪花，将水滴洒向空中。最后，辐射（电磁波、引力波、中微子）将从恒星流出，带走质量。

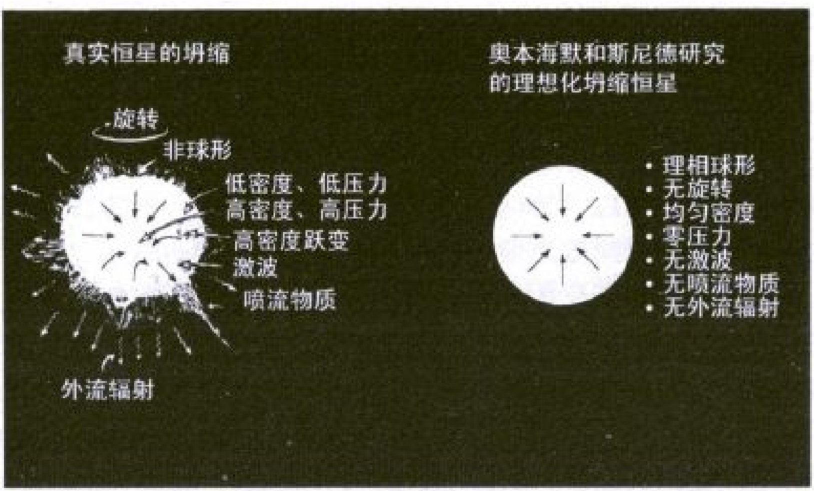



图6.3 左：发生在真实坍缩恒星的物理现象。右：奥本海默和斯尼德为计算星体坍缩而假定的理想情况。

奥本海默和斯尼德很想在他们的计算中包容所有这些效应，但这是不可能的，这样的计算远远超越了1939年的任何一个物理学家或计

算机器的能力。要等到80年代出现超大型计算机，那才能实现。所以，为了一点儿进步，必须建立坍缩恒星的理想化模型，然后为这个模型计算物理学定律的预言。

这样的理想化奥本海默太熟悉了：每当面对这么纷繁复杂的情况时，他几乎总能没有错误地分辨出哪些现象是核心的，哪些现象是边缘的。

对坍缩的恒星来说，奥本海默相信，有一个特征比其他任何特征更关键，那就是爱因斯坦广义相对论定律所描述的引力。在建立可行的计算时，它而且只有它是不能被破坏的。相反，恒星的自转和非球对称形状是可以忽略的，它们对某些坍缩的恒星可能有决定性的意义，但对缓慢旋转的恒星来说，它们可能不会产生强烈的影响。奥本海默不能从数学上证明，但从直觉看那是很清楚的，而且实际上也是正确的。同样，直觉告诉他，辐射的外流与激波和密度堆积一样，都是不太重要的细节。另外，由于（如奥本海默和沃尔科夫证明的那样）在大质量死星里，引力会超过一切压力，那么，似乎可以安然地假定（当然，这是错误的），坍缩的恒星没有任何内部压力——不论是热压力，还是电子和中子的幽闭简并运动产生的压力，或者核力产生的压力。一颗真实的恒星因为实际上存在着压力，所以它的坍缩方式可能会与理想化的没有压力的恒星不同；不过，奥本海默凭直觉认为，差别应该很小，不会太大。

这样，奥本海默与斯尼德提出一个理想化的计算问题：用广义相对论的精确定律研究一个完全球形、无自转、无辐射、密度均匀（表面附近的密度与中心的密度相同）、无任何内部压力的理想化恒星的坍缩。见图6.3。

即使做了这么多的理想化——在未来的30年里，其他物理学家总在怀疑这些理想化条件——计算还是极其艰难的。幸运的是，托尔曼就在帕萨迪纳，他能提供帮助。靠着托尔曼和奥本海默的指导，斯尼

德发现了决定整个坍缩的方程——他想了一个巧妙的办法来解决它们。现在，坍缩的所有细节他都能用公式表达了！从不同的角度审视这些方程，物理学家能读到他们感兴趣的关于坍缩的方方面面——从恒星外面看，它像什么？从里面看，它像什么？从恒星表面看，它又像什么？等等。[7]

坍缩恒星特别有趣的表现是从外面的静止参照系看到的。也就是说，观察者处在恒星外面的一个固定参照系中，而不随恒星塌缩的物质向内运动。从这个外面的静止参照系看，恒星开始坍缩的方式正是我们所料想的。像从屋顶下落的石头那样，恒星的表面起先也是慢慢下落（向内收缩），然后越落越快。如果牛顿的引力定律是正确的，坍缩就将持续加速，直到没有任何内部压力的恒星被高速挤压成一个点。据奥本海默和斯尼德的相对论公式，情形就不是这样的了。当恒星接近临界周长时，收缩会慢下来。随着恒星越来越小，它的坍缩也越来越慢，最后完全冻结在临界周长。不论等待多长时间，恒星外的静止观察者（即在外面的静止参考系中的人）都不可能看到恒星坍缩经过临界周长。这是奥本海默和斯尼德公式不容争辩的论断。

坍缩的冻结是不是由恒星内部某些未曾预料的广义相对论的力量引起的呢？不，奥本海默和斯尼德认识到，完全不是。实际上，它是由临界周长附近引力的时间膨胀（时间流慢了）引起的。在外面的静止观察者看来，坍缩恒星表面的时间在接近临界周长的过程中必然会越流越慢，相应地，发生在恒星表面和内部的各个事件，包括坍缩，都必然表现出缓慢的运动，然后逐渐冻结起来。

这件事情看来够特别了，而奥本海默和斯尼德的公式还预言了更特别的事情：虽然在外面的静止观察者看见坍缩在临界周长冻结了，但在恒星表面下落的观察者却会发现，坍缩根本没有冻结。如果恒星有几个太阳质量，从太阳大小开始坍缩，那么，从它自己表面上看，大约在1个小时内恒星就坍缩到了临界周长，然后继续坍缩下去，经过临界值，到达更小的周长。

到1939年，当奥本海默和斯尼德发现这些事情时，物理学家们已经习惯了时间是相对的事实；在宇宙中以不同方式运动的参照系所测量的时间流是不同的。但以前还没有谁遇到过不同参照系之间的如此极端的差别。从外面的静止参照系观测，坍缩将永远冻结，而在恒星表面的参照系中观测，坍缩将迅速地通过冻结点。研究奥本海默和斯尼德数学的人，没有谁会对如此极端的时间卷曲感到满意。不过，他们的公式就是这样的。人们可能欢迎启发性的解释，但似乎没有一个解释令人满意。到了50年代后期（本章结束时），我们才完全弄明白。

从恒星表面的观察者的观点看奥本海默和斯尼德的公式，我们甚至可以导出恒星沉没到临界周长以后的收缩情况；就是说，我们可以发现恒星收缩到无限大密度和零体积，而且可以导出在这一点的时空曲率的细节。然而，奥本海默和斯尼德在描述他们计算的文章里却回避了任何关于零体积的讨论。奥本海默在科学上天生是保守的，他不愿猜想（见第5章最后两段），这大概也令他不愿去讨论这些东西。

从公式中认识致密的零体积状态，这对奥本海默和斯尼德来说是太困难了；在1939年，即使临界周长和它外面的细节，对大多数物理学家来说也太离奇了。例如，在加州理工学院，托尔曼是相信的，毕竟这个预言是广义相对论不容争辩的结果。但别的人就不太相信。[8] 广义相对论只在太阳系内经受过实验检验，而那里的引力太弱了，牛顿定律差不多也能给出相同的预言。反过来，奥本海默和斯尼德的奇异预言依赖于超强的引力，多数物理学家想，在这么强大的引力面前，广义相对论也同样可能是失败的，即使没有失败，奥本海默和斯尼德也可能错误解释了他们的数学意义；即使他们没有错误解释，他们的计算也太过理想化，忽略了自转、密度堆积、激波和辐射，也是不能太当真的。

在美国和西欧，处处是这样的怀疑，但在苏联却没有。那时，朗道还在出狱后的恢复中，他有张“黄金名单”，记着发表在世界各地的

重要物理学研究论文。读到奥本海默和斯尼德的论文时，朗道也将它列入那个名单。他向朋友和伙伴们宣布，奥本海默的那些最新发现一定是正确的，虽然人类理解起来还极其困难。[9]朗道的影响是巨大的，他的观点从那时起就在苏联主要物理学家中间生根了。

## 插曲：原子弹

奥本海默和斯尼德是对还是错？假如没有第二次世界大战和后来发展氢弹的紧急计划，答案本来在40年代就该确切地知道了，但战争和氢弹确实来了，像黑洞那样不切实际的神秘研究停止了，物理学家的全部精力都转到了武器的设计。

到50年代后期，紧张的武器奋战才轻松了一些，星体坍缩又回到物理学家的意识中来。这个时候，怀疑者们向奥本海默一斯尼德预言发动了第一次严厉的攻击。首先高举怀疑大旗的是惠勒，但他没举多久；从一开始信仰者们的领袖就是惠勒的苏联对手，泽尔多维奇。

惠勒和泽尔多维奇的性格是在近20年的核武器计划的熔炉中形成的，那是40年代和50年代的20年，是黑洞研究冻结的20年。他们从核武器研究中走出来，带来了分析黑洞的重要工具：强有力的计算技术，对物理学定律的深刻理解，以及相互协作的研究作风，他们将以这种作风不断激发年轻的同事。他们也带来了沉重的包袱——与几个主要同行的一系列复杂的人际关系：惠勒与奥本海默，泽尔多维奇与朗道和萨哈洛夫（Andrei Dmitrievich Sakharov）。

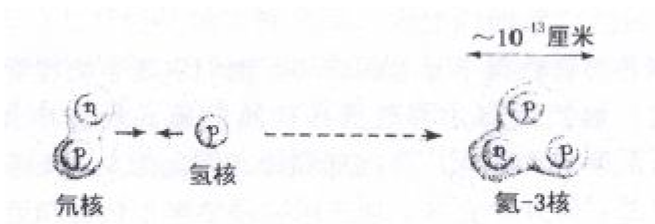
惠勒1933年刚研究生毕业，就获得了洛克菲勒基金资助的国家研究委员会（NRC）的博士后奖学金。他要选择到哪儿去、跟谁做博士后研究。他本可以像那时大多数NRC的理论物理学博士后一样去伯克利跟奥本海默，但他选择了纽约大学和布雷特（Gregory Breit）。惠勒说，“他们[奥本海默和布雷特]在个性上完全不同，奥本海默看事情黑白分明，绝对果断；而布雷特的工作总带着灰蒙蒙的色彩。我追求的东西需要长久的思考，所以我选择了布雷特。”[10]

1933年，惠勒从纽约大学去哥本哈根追随玻尔，然后到北加利福尼亚大学做副教授，接着又来到新泽西的普林斯顿大学。1939年，当

加利福尼亚的奥本海默和他的学生正在探索中子星和黑洞时，惠勒和玻尔正在普林斯顿（玻尔来访）发展核裂变的理论：[11]像铀那样的重原子核在受到中子轰击时会分裂成较小的碎片（卡片6.1）。那时，裂变刚被德国的哈恩（Otto Hahn）和斯特拉斯曼（Fritz Strassman）意外发现，而它的意义却是不祥的：通过裂变的一连串反应，可以造出威力空前的武器。但玻尔和惠勒并没考虑什么链式反应或武器，他们只想知道，裂变是怎么产生的？背后的机制是什么？物理学定律如何产生这个机制？

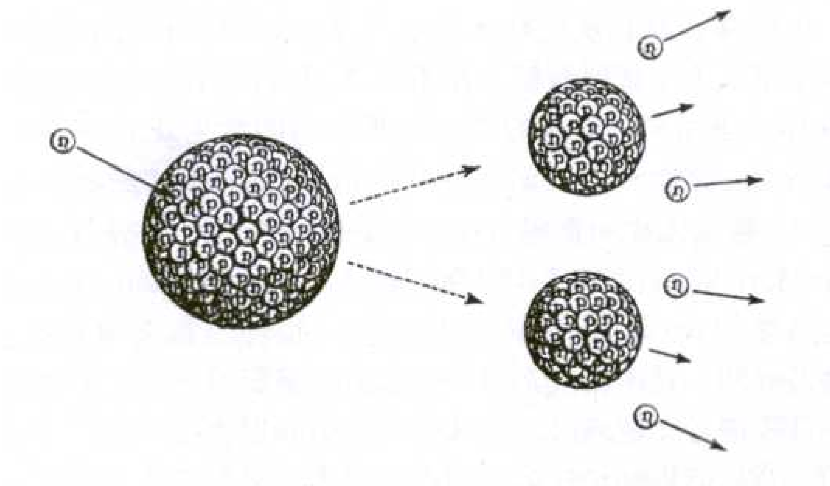
卡片6.1　聚变、裂变和链式反应

　　很轻的原子核聚变而形成中间大小的核，放出巨大能量。卡片5.3[22]的一个简单例子是1个，氘核（“重氢”，含一个质子和一个中子）和1个普通氢核（单个质子）形成1个氦-3核（两个质子和一个中子）的聚变：




　　这种聚变反应维持者太阳的热量，也是氢弹（在四五十年代，人们称它“超弹”）的能源。

　　很重的原子核裂变（分裂）而形成两个中间大小的核，放出巨大能量——远远超过化学反应的能量（因为决定核的核力远比决定化学反应原子的电磁力强大），但又远远小于轻核聚变的能量。有几种很重的核会自然裂变，而不需要任何外来帮助。本章更感兴趣的裂变是，中子轰击像铀-235（有235个质子和中子的铀核）那样的重核，将它一分为二：



有两个特殊的重核，铀-235和钚-239，它们的裂变不仅产生两个中间大小的核，还产生一群中子（如上面图中画的）。这些中子可能引起链式反应：假如铀-235和钚-239聚集在一个很小的包裹里，那么从一次裂变中放出的中子会撞击其他铀或钚的核，诱发它们裂变，生成更多的中子，而这些中子又诱发更多的裂变，如此下去。这种链式反应如果不受限制，将产生巨大的爆炸（原子弹爆炸），如果限制在反应堆里，那将是高效的电能。

玻尔和惠勒获得了显著的成功。他们发现了物理学定律如何产生裂变，他们预言了哪些核在持续的链式反应中最为有效，铀-235（后来成为毁灭广岛的那颗原子弹的燃料）和钚-239（这类核在自然界里并不存在，但美国物理学家很快会知道如何在核反应堆中产生，后来它点燃了毁灭长崎的那颗原子弹）。然而，玻尔和惠勒在1939年并没想原子弹，他们只想去理解。

玻尔和惠勒解释核裂变的文章，与奥本海默和斯尼德描述恒星坍缩的文章，发表在同一期《物理学评论》上，[11]发表日期是1939年9月1日，这一天，希特勒的军队入侵波兰，第二次世界大战爆发了。

1914年，泽尔多维奇出生在明斯克的一个犹太家庭，那年下半年，全家迁到圣彼德堡（20年代改称列宁格勒，90年代又恢复旧名）。泽尔多维奇15岁读完中学，后来没进大学，而是到列宁格勒的物理技术研究所的实验室当助手。他在那儿自学了很多物理和化学，也做了许多动人的研究，结果，没有受过任何正规大学训练的他在1934年获得了博士学位，那年他20岁。

1939年，当惠勒和玻尔正在探索核裂变理论时，泽尔多维奇和他的一个亲密朋友哈里顿（Yuli Borisovich Khariton）正在研究核裂变产生的链式反应的理论。激发他们研究的是法国物理学家佩兰（Francis Perrin）的一个有趣（但是错误）的设想：火山喷发的动力可能来自天然的地下核爆炸，那是原子核裂变链式反应的结果。不过，没人（包括佩兰）知道这种链式反应的细节。泽尔多维奇和哈里顿——已经走进世界最优秀的化学爆炸专家的行列——向着这个问题冲锋了。几个月里，他们（同时西方也有人）证明了这类爆炸不可能自然发生，因为自然出现的铀大多是铀-238，而没有足够的铀-235。不过，他们下结论说，如果人工分离出铀-235，将它们聚集起来，那就可以制造链式反应爆炸。（美国人马上就会开始这种分离，为他们的“广岛原子弹”制造燃料。）罩着核研究的秘密帷幕还没有降下，泽尔多维奇和哈里顿将他们的计算发表在最有名的苏联物理学刊物《实验与理论物理学杂志》上，让全世界都来看。[12]

在6年的战争中，参战国的物理学家们发明了声纳、扫雷器、火箭、雷达以及最要命的原子弹。在新墨西哥州洛斯阿莫斯，奥本海默领导“曼哈顿计划”，设计和制造美国的原子弹；在华盛顿汉佛，惠勒作为首席科学家，设计和建造世界上第一个生产规模的原子核反应堆，为“长崎原子弹”制造燃料。[13]

广岛和长崎在原子弹中毁灭，几十万人在原子弹下丧生，令奥本海默很痛苦：“如果原子弹一定要添加进战争国家的军火库或者准备战争的国家的军火库，那么人类总会有诅咒洛斯阿莫斯和广岛的一

天。”[14]“粗俗也罢，幽默也罢，夸张也罢，总之从某种意义说，物理学家已经认识到犯罪了；而这是他们不能失去的良知。”[15]

惠勒却另有一样悔恨：“回头来看[1939年我与玻尔关于裂变理论的研究]，我感到很难过。我首先是以物理学家的眼光来看核裂变的[那只不过是为了好奇，想知道裂变是怎么回事]，其次才以一个公民的眼光来看它[为了我的国家的国防]，怎么会是这样呢？为什么我没想到自己首先是一个公民，其次才是物理学家呢？简单看看记录就知道，约2 000万到2 500万人在二战中死亡，而大多数是在后几年死的。战争每缩短一个月，就可能拯救50万到100万的生命，我的哥哥乔也可能会活下来，他是1944年10月在意大利的战斗中死去的。假如[第一次在战争中使用原子弹]那转折的一天不是1945年8月6日，而是1943年8月6日，该会有多大的不同啊！”[16]

在苏联，物理学家在1941年6月德国来进攻时放弃了所有的核研究，因为对国防来说别的物理学领域能更快产生结果。德军压进，包围列宁格勒时，泽尔多维奇和朋友哈里顿撤退到卡赞，在那儿热烈地研究常规炸弹的理论，以提高它们的威力。后来，他们在1943年被召到了莫斯科。他们听说，美国和德国正加紧制造原子弹，这是很明显的。他们将加入库尔恰托夫（Igor V.Kurchatov）领导的一个杰出小组，发展苏联的原子弹。

两年后，美国原子弹在广岛和长崎爆炸时，库尔恰托夫的小组彻底从理论上认识了生产钚-239的核反应堆，设计了几种可能的原子弹——哈里顿和泽尔多维奇成了这个计划的理论领导。

斯大林知道美国原子弹爆炸后非常生气，大骂库尔恰托夫，恨他的苏联小组进展太慢。库尔恰托夫为他的小组辩解说，在战争蹂躏下，凭有限的资源，他的小组的进展已经够快了。斯大林愤怒地告诉他，如果小孩儿不哭，妈妈怎么知道他想要什么？！你需要什么就说，可以答应一切要求！然后，他又要求实施一个行动不受限制的原

子弹紧急计划，由那位可怕的秘密警察领导人别里亚（Lavrenty Pavlovich Beria）全权负责。

很难想象别里亚费了多大的力气。斯大林监狱里的几百万人被迫应征，这些“泽克”们（在口头上人们这样称呼）[1]建立了铀矿、铀提炼厂、核反应堆、理论物理研究中心、武器试验中心和为这些机构提供自给服务的小城市。它们散布在全国各地，迷雾重重，为的是不让美国的曼哈顿计划知道。泽尔多维奇和哈里顿被调往其中的一处，在“一个遥远的地方”，它的位置，尽管西方国家在50年代末就几乎确定了，但苏联到90年代才向国人公开。[2]他们去的地方简称是Obyeht（“基地”），哈里顿是负责人，而泽尔多维奇负责下面一个关键的原子弹设计小组。在别里亚领导下，库尔恰托夫建立了几个物理学家小组，同时而完全独立地探讨原子弹计划的各个方面；尽管烦琐复杂，但能保证安全。基地的小组向其他小组提出设计问题，其中一个小组的领导者是莫斯科物理问题研究所的朗道。

当研究队伍滚滚向前时，苏联间谍通过福奇斯（Klaus Fuchs，参加美国原子弹计划的英国物理学家）获得了美国钚弹的设计，它与泽尔多维奇和他的同事们提出的设计有些不同。这样，库尔恰托夫、哈里顿和伙伴们就面临着艰难的抉择：他们承受着沉重的压力，斯大林和别里亚在等着结果；他们害怕原子弹试验失败的后果，在那个年代，失败就意味着杀头。他们知道美国的设计在广岛和长崎成功了，而他们有的钚只够一颗原子弹。如何抉择是显然的，也是痛苦的：他们将自己的设计放到一边，[3]将他们的紧急计划照着美国的设计进行修改。[17]

4年过去了，经过无数紧张的奋斗，不知有多少秘密，说不清死了多少被劳役的犯人，在舍里亚宾斯克附近的核反应堆产生了多少废物——10年后那儿发生爆炸，污染了几百平方英里的土地[18]——紧急计划在1949年8月29日终于到头了。第一颗苏联原子弹在亚洲的谢米帕拉京斯克附近爆炸了，苏军最高司令和政府官员现场观看了试验。

1949年9月3日，一架美国WB-29气象探测飞机在从日本飞往阿拉斯加的航线上，发现了苏联核裂变试验的产物，数据被送给包括奥本海默在内的一个专家委员会评估，毫无疑问地证实了苏联试验了一颗原子弹！

这给美国带来了恐慌（如后方的原子弹隐蔽所，学校孩子们的原子弹演习，麦卡锡在政府、军队、媒体和大学中清理共产党间谍和他们的同党的政治迫害），在物理学家和政治家之间爆发了巨大的争论。在美国原子弹设计中最具创造力的物理学家特勒（Edward Teller）倡导尽快设计和建造“超弹”（或“氢弹”）——一种以氢核聚变形成氦为基础的武器。如果造出氢弹来，那将是很可怕的。似乎没有办法限制它的威力。谁想过比广岛原子弹威力大十倍、百倍、千倍甚至百万倍的炸弹吗？如果氢弹成功了，它的威力要多大有多大。

惠勒支持特勒。他相信为对抗苏联的威胁，“超弹”的紧急计划是基本的。奥本海默和他的美国原子能委员会一般咨询委员会表示反对。他们指出，人们相信的“超弹”是否能建造成功，还并不是很清楚。退一步讲，即使成功了，那些威力比普通原子弹大得多的超弹也可能会太重，不能用飞机或火箭来发射。另外，还有道德上的问题。奥本海默和他的委员会说，“我们相信，这个建议[紧急计划]必然会带来的对人类的极大危险完全超过了可能由它带来的军事上的优势，这也是我们反对它的原因。我们要让大家都清楚地认识到，这是一种超级武器，与原子弹属于完全不同的范畴。要发展这种超弹，就得容忍用一颗炸弹去毁灭一大片；使用这种武器，意味着决心去屠杀大量的国民。至于几个可能规模的超弹的爆炸会产生多大的全球性的辐射影响，我们感到更可怕了。假如超弹成功了，它的破坏力是没有极限的；因此，超弹可能成为种族灭绝的凶器。”[19]

在特勒和惠勒看来，这些理由是毫无意义的。他们相信，俄国人肯定会向氢弹推进的，如果美国落后了，自由世界就将陷入巨大的危险。

特勒和惠勒的观点占了上风。1950年3月10日，杜鲁门总统下令实施超弹紧急计划。

现在看来，美国1949年的超弹计划似乎真像奥本海默的委员会猜测的那样，是一个失败的策略。不过，它不是一定会失败，而当时对它也没有更好的认识，所以大家都还在努力探索，到1951年3月，特勒和乌拉姆（Stanislaw Ulam）发明了一种全新设计，看来大有希望。

特勒和乌拉姆发现的只是一种设计思想。正如贝特（Hans Bethe）所说，“特勒的想法，十个有九个是没用的。他需要判断力更强的人，即使天赋差些，来帮他选择第十个思想，那常常是天才的灵感。”[20]为了检验这个思想是天才的灵感还是骗人的废物，需要将它落实到具体周详的原子弹设计中去，然后用最大的计算机来执行那些复杂的计算，看设计是否可行；如果计算成功，那么接下来就是制造和试验真正的原子弹了。



1952年，普林斯顿大学惠勒氢弹设计小组部分成员。左起，前排：Margaret Fellows,Margaret Murray,Dorothea Ruffel,Audrey Ojala,Christene Shack,Roberta Casey。第二排：Walter Aron,William Clendenin,Solomon Bochner,John Toll，惠勒，Kenneth Ford。第三、四排：David Layzer,Lawrence Wilets,David Carter,Edward Frieman,Jay Berger,John McIntosh,Ralph Pennington，不详，Robert Goerss。

[Howard Schrader摄；Lawrence Wilets和惠勒提供。]

为计算设立了两个小组：一个在洛斯阿莫斯，另一个在普林斯顿大学。惠勒负责普林斯顿那组。惠勒的小组夜以继日几个月，在特勒-乌拉姆思想基础上提出了一个完整的原子弹设计方案，并用计算机检验了它是否可行。惠勒后来回忆，“我们做了大量的计算，用了纽约、费城和华盛顿的计算机设备——实际上，全美国大部分的计算机都用上了。为了弄清问题，维勒茨（Larry Wilets）、托尔（John Toll）、福德（Ken Ford）、亨尼耶（Louis Henyey）、豪斯曼（Carl Hausman）、奥利维尔（Dickl'Olivier）等人每天6个小时三班倒。”[21]

计算清楚说明，特勒-乌拉姆思想大概是可行的。在普林斯顿高等研究院（奥本海默是那儿的院长）召开了一个会，向奥本海默的一般咨询委员会和它的上级美国原子能委员会报告这个思想。特勒先作介绍，然后，惠勒讲述了他的小组的具体设计和得到的爆炸结果。惠勒后来回忆说，“在我开始讲话时，福德从外面赶到窗口，将窗户抬起，递进一大卷图。我将图展开，挂在墙上，它说明了[我们所计算的]热核燃烧的过程……委员会没有别的选择，只得承认这些事情还是有意义的……我们的计算也让奥比改变了对这个计划的看法。”[22]

奥本海默也讲过他自己的感受：“我们在1949年的计划[那个‘失败的策略’]是一个被扭曲了的东西，你可以有充分理由认为它不会有多大技术意义，于是还可能认为，即使有了它，也并不需要。1952年的计划[以特勒-乌拉姆思想为基础的新设计]在技术上是很吸引人的，没有什么可以争论的。问题只是军事的、政治的和人道的，那就是一旦你有了超弹，你要怎么做？”[23]他没有再表示对有关伦理学问题的忧虑，同委员会其他成员和特勒、惠勒以及超弹的拥护者们走到了一起；制造和试验超弹的计划在加速向前，像惠勒小组和洛斯阿莫斯小组同时进行的计算所预料的那样，进展很顺利。

惠勒小组复杂的设计计算最后写成秘密的《马特哈恩计划，B部，报告31》或PMB-31。惠勒说，“我听说，至少在10年里，PMB-31是热核装置[氢弹]设计的圣经。”[24]

1949～1950年，当奥本海默、特勒等人在美国的恐慌中争论是否该紧急发展超弹时，苏联已经在施行他们自己的超弹计划了。

1948年春，苏联第一颗原子弹试验15个月前，[25]泽尔多维奇和他的小组在“基地”完成了类似于美国“失败策略”的关于超弹设计的计算。[4]1948年6月，在苏联最著名的理论物理学家之一的塔姆（Igor Tamm）领导下在莫斯科建立了第二个超弹小组。[26]成员有金兹堡（Vitaly Lazarevich Ginzburg，在第8、10章我们会听到他更多的事

情）、萨哈洛夫（70年代他将成为一个异议人士，到80和90年代，他又将成为苏联的英雄和圣人）、别连斯基（Semyon Belenky）和罗玛诺夫（Yuri Romanov）。塔姆小组的任务是审查和优化泽尔多维奇小组的设计计算。

他们对这个任务的态度，还是别连斯基说得干脆："我们的工作就是为泽尔多维奇舔屁股。"[28]泽尔多维奇的个性很矛盾，在学术上坚强有力，对别人要求苛刻，而政治上胆小懦弱，在苏联物理学家中不大讨人喜欢，但他的确是很卓越的。朗道领导着下面的一个辅助设计小组，偶尔会接到泽尔多维奇的命令，要他就原子弹的设计分析这分析那。有时，他会在背后说他是"母狗，泽尔多维奇"。[29]不过，泽尔多维奇倒是很尊重朗道，认为他是判决物理学思想的伟大法官，是他最伟大的老师——虽然他从来没有正式听过朗道的一堂课。

塔姆小组的萨哈洛夫和金兹堡只用几个月就得到一个远比泽尔多维奇和美国人正在探寻的"失败的策略"好的超弹设计方案。萨哈洛夫提出造一个重裂变层（铀）和轻聚变层相间的千层饼式的超弹，金兹堡提出轻聚变燃料用氘化锂（LiD）。[30]在超弹强烈爆炸中，LiD的锂核将裂变成氚核，氚核将与氘核一起聚合成氦核，释放出巨大能量。重元素铀会阻止能量过快释放，帮助压缩聚合燃料，向燃料加入裂变能量，通过这些过程，它会增加爆炸的威力。当萨哈洛夫报告这些思想时，泽尔多维奇一下子就看到了希望。萨哈洛夫的千层饼和金兹堡的LiD很快就成为苏联超弹计划的焦点。

为将超弹计划更快向前推进，萨哈洛夫、塔姆、别连斯基和罗玛诺夫奉命从莫斯科迁往基地，但金兹堡没去，原因是很清楚的：3年前他与伊凡诺夫娜（Nina Ivanovna）结婚，她是一个活泼而有才华的女人，40年代初被人诬陷谋害斯大林而进了监狱。有人告她和她的伙伴密谋从她家的窗户向下面经过阿尔巴特街的斯大林开枪。正当三人法官小组判决时，有人指出她家的窗户根本看不到阿尔巴特街。于是，她享受了难得的恩赐，命保住了，只被判坐牢，然后被流放，没

有死。夫人的坐牢流放大概足以玷污金兹堡的名誉，这位超弹的LiD燃料的发明者只好被基地拒之门外。他高兴了，比起超弹设计，他更喜欢基础的物理学研究；而科学世界也得到了他的回报：在泽尔多维奇、萨哈洛夫和惠勒集中精力研究超弹时，金兹堡揭开了宇宙线如何在银河系中传播的秘密，又同朗道一起用量子力学定律解释了超导性的起源。

1949年，当苏联原子弹计划结束时，斯大林命令动员全国一切力量立即投人超弹计划。从囚犯到理论研究人员、工厂、试验场以及有关设计和建造各方面的物理学家小组都要集中起来，在氢弹上与美国斗争。而美国人对此却一无所知，他们还在争吵是否需要赶紧发展超弹。不过，美国技术领先，起步也快。

1952年11月1日，美国爆炸了一个代号麦克（Mike）的氢弹型的装置。麦克是为了检验1951年的特勒-乌拉姆发明而设计的，基础是惠勒和洛斯阿莫斯两个小组同时进行的设计计算。它以液氘为主要燃料。为了液化氘并注入爆炸区，需要工厂那么大的设备，所以这不是那种可以用任何飞机或火箭发射的炸弹。不管怎样，它还是完全摧毁了西太平洋埃尼威托克岛的埃卢格兰礁屿，威力比在广岛令100 000人丧生的那颗原子弹大800倍。[31]

1953年3月5日，莫斯科广播电台在哀乐声中宣布约瑟夫·斯大林逝世。美国人高兴了，苏联一片伤悲。萨哈洛夫给妻子兑拉娃写信说，“伟人的死给我带来了巨大的影响，我正在想他这个人。”[32]

1953年8月12日，在舍米帕拉京斯克，苏联爆炸了他们的第一颗氢弹，美国人称它为“乔-4”（Joe-4），是照萨哈洛夫的千层饼设计的，聚变燃料用金兹堡的LiD，体积很小，可以用飞机投放。不过，乔-4的燃料不能用特勒-乌拉姆方法点燃，结果它的威力就远比美国的麦克小：它“只有”广岛原子弹的30倍，而麦克是800倍。

实际上，用美国超弹设计者的话说，乔-4根本不是氢弹，而是加强的原子弹，也就是说，在原子弹中加入了一些聚变燃料，从而增大了威力。这种加强的原子弹早就在美国的军火库里了；美国人不承认它们是氢弹，因为千层饼的设计方法不能点燃任意量的聚合燃料。举例说，凭这种设计没办法造出比广岛原子弹强几千倍的“世界末日武器”。

但30颗广岛原子弹也是不能轻视的，何况它还可以用飞机投放。不过，乔-4实际上只是令人敬畏的武器，而惠勒和其他美国人却长长舒了口气，幸好自己有真正的氢弹，苏联新领导人马林科夫（Georgi Malenkov）不能用它来威胁美国了。

1954年3月1日，美国爆炸了第一颗LiD燃料的可投放氢弹，代号为布拉沃（Bravo），同麦克一样，它的基础是惠勒和洛斯阿莫斯小组的设计计算，也用特勒-乌拉姆的发明。爆炸能量是广岛原子弹的1 300倍。

1954年3月，萨哈洛夫和泽尔多维奇（独立于美国）联合发现了特勒-乌拉姆思想。[33]几个月内，苏联的力量都集中来实现一颗真正的超弹，破坏力要多大有多大。他们只用了18个月就完成了设计和建造，1955年11月23日，它突然爆炸了，能量是“广岛”的300倍。

正如奥本海默的一般咨询委员会在反对超弹紧急计划时所猜测的那样，这些大能量的超弹——以及后来苏联为了恐吓肯尼迪而爆炸的5000个广岛原子弹威力的庞然大物——并没引起美国和苏联军队的多大兴趣。现在俄国和美国军火库里有的只是约30个广岛的原子弹，没有几千倍的。尽管它们是真正的氢弹，却并不比大原子弹更有威力。军队不需要也不想要“世界末日”的东西。这种东西惟一的作用在于给敌人以心理上的威慑——在有像斯大林那样的领导者的世界里，这种威慑作用还是很重要的。

1953年7月21日，斯特劳斯（LeWiS Strauss）成为美国原子能委员会主席，他当委员时曾在超弹紧急计划问题上与奥本海默激烈争论过。他当权的头一件事就是将普林斯顿的奥本海默办公室的所有秘密材料拿走。斯特劳斯和华盛顿的许多人都很怀疑奥本海默的忠诚。一个忠于美国的人怎么可能像他那样在惠勒小组证明特勒-乌拉姆思想前反对超弹计划呢？原子能国会联合委员会的首席律师波尔登（William Borden）在超弹争论期间给胡佛（J.Edgar Hover）写信说，“此信的目的在于表达我个人深思熟虑的意见。通过对所有秘密证据的多年研究，J·罗伯特·奥本海默很有可能是苏联的代理人。”奥本海默的忠诚审查被取消了，1954年4、5月，在美国第一次投放氢弹试验期间，原子能委员会举行听证会，决记奥本海默是否真的危及同家安全。

听证会期间，惠勒正在华盛顿忙别的事情，没有卷进来。然而，他的亲密伙伴特勒在去作证的前一天晚上来到他所在宾馆的房间，在地板上走来走去，彷徨了几个小时。假如特勒说他想说的话，会极大伤害奥本海默；但他能不那么说吗？惠勒没什么疑虑，在他看来，特勒很正直，一定会说实话。

惠勒是对的。第二天，特勒说的话大概只有惠勒能理解：“在很多时候，我看到奥本海默博上做的事情……我非常难以理解。在许多观点上，我完全不同意他的，而他的行为对我来说太复杂了，简直令我疑惑。在这一点上，我感觉应该考虑目前最重要的国家利益，这一点我认识更多，也相信更多……我相信——当然，这只是信仰问题，没有什么专业问题和背景——奥本海默博士，凭他的性格，是不会故意也不会愿意做危害国家安全的事情的。所以，如果你们的问题说的是一种倾向的话，我可以说，我找不到什么理由说他不忠；如果问题是要凭他在1945年以来的行为来作出明智的判决，那么我可以说最好也不要肯定他的忠诚。”[34]

几乎所有来作证的物理学家都无疑是支持奥本海默的——特勒的

证词令他们惊呆了。尽管如此，尽管没有可靠的证据说明奥本海默是“苏联的代理人”，时代的潮流还是占了上风；奥本海默被宣布危害国家安全，是对国家不忠的人，没有资格参与秘密事务。

在多数美国物理学家看来，奥本海默是时代的牺牲品，而特勒是小人。特勒在他的余生将被物理学界唾弃。但在惠勒看来，特勒才是牺牲者，他相信，特勒“凭勇气表达了他忠实的判断，他把国家利益放在物理学家的团结之上。”[35]惠勒认为，他的证词“值得考虑”而不该被排斥。35年后，萨哈洛夫也同意这一点。[5][36]

[1] zek，苏联监狱称犯人的黑话，是zaklyuchennyi监狱的简称。——译者

[2] 在舍里亚宾斯克和乌拉尔山之间的阿兹马斯城附近。

[3] 在美国设计基础上的原子弹试验成功以后，苏联人又回到自己的设计，在它基础上建造了原子弹，1951年试验成功。

[4] 萨哈洛夫猜测，这个设计可能直接受通过间谍活动（也许通过间谍福奇斯）从美国得到的情报的激发。而泽尔多维奇则断言，不论福奇斯还是别的间谍，都没有提供过他的设计小组所不知道的任何有价值的情报；苏联超弹间谍活动的最大作用在于让苏联政治当局相信他们的物理学家知道自己在做什么。[27]

[5] 在这一点上，我决不赞同惠勒（尽管他是我最亲密的朋友和老师）和萨哈洛夫。关于特勒-奥本海默的矛盾和美国是否该造原子弹的争论，我推荐大家去看Bethe（1982）和York（1976）的思想性和知识性都很强的著作。关于萨哈洛夫的观点，可以看他1990年的东西；贝特（1990）对他有过批评。奥本海默听证会的实况，可以看美国原子能委员会（USAEC）1954年的记录。

# 黑洞的诞生：深入理解

惠勒和奥本海默不仅在国家安全问题上有重大分歧，他们的理论物理学方法也大相径庭。奥本海默只信奉已经建立起来的物理学定律的预言，而惠勒渴望知道那些定律背后的事情。他的思想常常走近旧定律破灭而新定律出现的边缘。他想从他的道路跳进21世纪，去看看20世纪物理学前沿之外的那些物理学定律会是什么样子。

他可能看到些什么呢？从50年代起，在惠勒看来，未来最有希望的莫过于广义相对论（宏观领域）和量子力学（微观领域）的交流，那时两者还没有逻辑一致地融和起来。起初尝试解决这个问题时，它们就像一个纵横字谜。你可以试着用一组词横填，用另一组词纵填。你会发现纵横交叉处出现了矛盾：横的那行GENERAL需要E，而纵的那列QUANTUM需要U；横的RELATIVITY要E，而纵的QUANTUM要T。看看这些行列，就会明白，要么改一个字，要么两个都改，才会没有矛盾。类似地，对广义相对论和量子力学，为了从逻辑上将两者融和起来，显然应该修改其中的一个，或者两个都要修改。如果这种融和实现了，那么广义相对论与量子力学的最终统一将产生一组新的强有力的定律，物理学家称它为量子引力。不过，50年代的物理学家对两者结合的认识还太简单，尽管费力不少，但进展不多。




进展同样缓慢的还有对原子核的基本构成要素的认识——中子、电子以及粒子加速器正在产生的越来越多的基本粒子。

惠勒想跳过这个泥潭，一眼看到量子引力和基本粒子的本性。他认为，找出理论物理学中疑惑多的地方，就能看到这些本性。疑惑的背后就是深刻的认识。疑惑越深，我们的认识超越20世纪物理学前沿可能就越远。

惠勒刚从超弹中解脱出来，就带着这种精神，同哈里森和若野一起填补了我们对冷死星体认识的空白（第5章）；他也带着这种精神思考“大质量物体最终命运”的问题。他在这儿遇到了一个他正在寻找的疑惑：没有比2个太阳质量更大的冷死星；然而天空中却有很多质量远比它们大的热星——总有一天会冷却死亡的。奥本海默曾以一贯的直率作风向确立的物理学定律寻求那些星体的行为，（与斯尼德一起）得到一个答案，这似乎令惠勒感到奇怪，也更令惠勒相信，他可能会在这个大质量的命运问题上发现超越20世纪前沿的物理学。我们

将在第12、13章看到，惠勒是对的。

惠勒满怀火一样的热情——深深地、不懈地追寻大质量物体的命运，想知道它们的命运是否能解开量子引力和基本粒子的秘密。奥本海默不一样，他在1958年似乎对很多事情都不太关心。他相信自己和斯尼德的计算，但没想过把它更推进一步，也没兴趣更深入去认识它。也许他太累了，在过去的20年里，他一直在斗争——武器设计的斗争，政治的斗争，个人的斗争。也许他在未解之谜面前畏缩了。不管怎么说，他再也没有回答这些问题。火炬传给了新一代。奥本海默留下的东西，是惠勒的基础；而在苏联，朗道留下来的，也成了泽尔多维奇的基础。

1958年在布鲁塞尔遇到奥本海默时，惠勒断言奥本海默-斯尼德计算是靠不住的。为什么呢？因为他们的理想化条件太苛刻了（见图6.3）。最特别的是，奥本海默一开始就假定坍缩的恒星不存在任何压力。如果没有压力，坍缩的物质就不可能形成激波（类似于破碎的带着泡沫的海浪）。没有压力和激波，坍缩的物质就无法加热。没有热量和压力，就无法引起核反应，无法产生辐射。没有通过核反应、压力和激波产生的辐射外流和物质喷发，恒星就无法失去质量。一开始就禁止了质量的损失，大质量恒星就不可能将质量减到2个太阳以下而成为冷死的中子星。惠勒由此推论，奥本海默的坍缩星体无疑会形成黑洞；到头来，他的理想化什么事也做不成！

奥本海默和斯尼德在1939年做他们的研究时，还不可能具体计算具有真实压力（热压力、简并压力和核力产生的压力）的伴随着核反应、激波、热辐射和物质喷射的坍缩过程。不过，核武器设计的20年正好提供了必需的工具。压力、核反应、激波、热、辐射和质量喷射，都是氢弹的重要特性；没有它们，氢弹是不会爆炸的。为设计氢弹，需要将所有这些事情放到计算机里一起计算。当然，惠勒的小组就做这件事。所以，他们现在很自然地会重写原来的计算程序，不过不是去模拟氢弹的爆炸，而是来模拟大质量恒量的坍缩。

假如那个小组还在，这会是很自然的事情。然而，小组解散了。他们曾写过PMB-31报告，分散后，有的当了老师，有的搞物理学研究，还有的成了各大学和政府实验室的行政领导。

美国原子弹设计机构现在都集中在洛斯阿莫斯和加利福尼亚利弗莫尔的一个新的政府实验室里。50年代后期，科尔盖特（Stirling Colgate）在利弗莫尔迷上了星体坍缩问题。他在特勒鼓励下，与怀特（Richard White）、后来又与麦（Michael May）合作，在计算机上模拟坍缩。他们的模拟还保留了一些奥本海默的理想化：一开始就认定坍缩的恒星是无旋转球形的。如果不这样限定，计算会困难得多。不过，他们的模拟还是将困惑惠勒的那些东西都考虑进来了——压力、核反应、激波、热、辐射、质量喷射——他们所依靠的是原子弹设计的专业知识和计算机编码。完善设计费了几年的努力，但到60年代初，他们已经做得很好了。

60年代初的一天，惠勒急匆匆赶到普林斯顿大学的一个相对论讨论班，我那时正在那儿跟他读研究生。他来晚了一点儿，但满脸高兴。他刚从利弗莫尔访问回来，在那儿看到了科尔盖特、怀特和麦的最新模拟结果。他在黑板上画了一个又一个圆，用激动的声音向我们解释他的利弗莫尔朋友们知道了什么：

如果坍缩恒星的质量小，它将导致超新星爆发，像茨维基30年前猜想的那样形成中子星。当恒星质量超过2个太阳的最大中子星质量时，不管多大压力、也不管核反应、激波、热和辐射，它都将坍缩形成黑洞。[37]而黑洞的诞生显著地类似于奥本海默和斯尼德在近25年前高度理想化的计算。从外面看，坍缩会越来越慢，最后在临界周长处冻结；在星体表面的人看来，坍缩一点儿也没冻结。星体的表面会毫不犹豫地一直收缩下去，穿过临界周长。

其实，惠勒已经在期待这一点了。由于其他的认识（下面讲），他已经从奥本海默黑洞的批评者变成了热情的支持者。不过他开始相

信的还是一个具体的来自实际计算机模拟的证明：坍缩必然产生黑洞。

奥本海默为惠勒的转变高兴吗？不，他几乎没表现出什么兴趣和愉快。1963年12月，当类星体发现时（第9章），在德克萨斯的达拉斯举行了一个国际会议，惠勒在会上发表了关于星体坍缩的长篇演说，热情讲述了奥本海默和斯尼德1939年的计算。奥本海默也参加了会议，惠勒演讲时，他正坐在走廊的长椅上跟朋友聊别的事情。30年后，惠勒回想这一幕，说起来还很难过，两眼满是伤悲。

50年代后期，泽尔多维奇开始对核武器设计工作感到厌倦，大多数真正有意思的问题都已经解决了。为寻找新的挑战，他用一部分时间去尝试基本粒子的理论，然后又转向天体物理学。同时，他仍然领导着基地的一个原子弹设计小组和一个在莫斯科应用数学研究所里做辅助计算的小组。

在原子弹设计工作中，泽尔多维奇常用一些思想来激发他的小组，组员们会通过计算来看这些思想是否行得通。“泽尔多维奇点火，队员们加油”，这是金兹堡的说法。在天体物理学研究中，泽尔多维奇仍然保持着这种作风。

在天体物理学问题中，令泽尔多维奇感兴趣的是星体的坍缩。像美国的惠勒、科尔盖特、麦和怀特一样，泽尔多维奇也明显感到，氢弹设计的工具完全适用于坍缩星的数学模拟。

为弄清真实星体坍缩的具体情况，泽尔多维奇留了几个年轻人在身边：应用数学所的纳杰任（Dmitri Nadezhin），伊姆舍尼克（Vladimir Imsheimik）和基地的波杜利茨（Mikhal Podurets）。经过一系列激烈讨论，他向他们讲了自己如何考虑在计算机上模拟坍缩，并包括那些对氢弹来说十分重要的效应：压力、核反应、激波、热、辐射、物质喷射。

在这些讨论启发下，伊姆舍尼克和纳杰任模拟了小质量恒星的坍缩，还独立于美国的科尔盖特和怀特证实了茨维基关于超新星的猜想。同时，波杜利茨模拟了大质量恒星的坍缩，他的结果几乎与麦和怀特的同时发表，而且在内容上与那两位美国同行也差不多是一样的。[38]没有什么可以怀疑的了：坍缩产生黑洞，而且正是按照奥本海默和斯尼德声明的那种方式。

原子弹设计代码能用来模拟星体坍缩，这不过是核武器与天体物理学众多密切联系的一个表现。那些联系，萨哈洛夫在1948年就看清了。他奉命参加塔姆的原子弹设计小组时，就开始以天体物理学研究来为自己做准备。1969年，我也意外地卷进了这些联系。

我从没真正想过要明白特勒-乌拉姆和萨哈洛夫-泽尔多维奇的思想是什么。凭他们的思想，那些能“具有任意威力”的超弹，在我看来是很讨厌的东西，我甚至根本不去想它如何实现。但我想知道中子星在宇宙中会起什么作用，只好被迫去考虑特勒-乌拉姆思想。

泽尔多维奇几年前指出，星际空间或恒星附近的气体在落到中子星上时会加热发光，它辐射出来的实际上大多是高能的X射线，而不是低能的可见光，那该有多热！泽尔多维奇认为，下落的气体控制着X射线外流的速率；反过来，外流的X射线也影响着气体下落的速率。于是，协同作用的气体和X射线产生一种稳定的自调节流。如果下落气体太快，它将产生大量X射线，而外泻的X射线将与下落的气体相撞，产生向外的压力，从而减缓气体的下落（图6.4（a））。另一方面，如果气体下落太慢，它产生的X射线将很少，没有能力减缓气体的下落，从而下落速率会增大。只存在一个惟一的气体下落速率，不太高，也不太低，正好能让X射线和气体达到平衡。

气体与X射线的这幅流动图景令我困惑。我清楚地知道，在地球上，假如想用低密度液体（如水）在下面支撑高密度液体（如水银），那么水银会很快像舌头一样钻进水中，水银下沉，而水上升

（图6.4（b）），这个现象叫瑞利-泰勒不稳定性。在泽尔多维奇的图景中，X射线就像低密度的水，而下落的气体像高密度的水银。气体的舌头难道不会钻进X射线吗？然后，气体不会自由落向那些舌头，从而破坏泽尔多维奇的自调节流吗（图6.4（c））？用物理学定律详细计算，可能会告诉我这种事情是否发生，但这样的计算太复杂，又费时间；所以我没有算，而直接去问泽尔多维奇，那是1969年的一个下午，我们正在他的莫斯科公寓里讨论物理学。

我提出这个问题时，泽尔多维奇有点儿不高兴，但他的回答很坚决："不，基普，那是不会发生的。没有舌头钻进X射线，气流是稳定的。""您怎么知道呢，雅科夫波里索维奇？"我问。奇怪的是，我没能得到回答。显然，泽尔多维奇或其他什么人做过详细计算或实验，证明了X射线能顶住气体而不会出现瑞利-泰勒的舌头来破坏，但泽尔多维奇却不能向我指出发表在文献里的任何一家的计算或实验，也没有为我详细讲述其中的物理学。这太不像泽尔多维奇了！

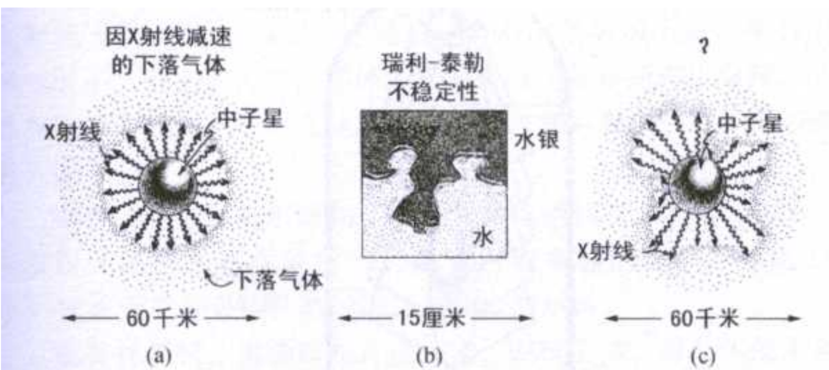



图6.4 （a）落向中子星的气体被外流的X射线压力减速。（b）水银在地球的引力作用下下落，它下面的水将它支撑着，结果出现瑞利-泰勒不稳定性。（c）下落的气体被中子星的X射线所支撑，也会出现瑞利-泰勒不稳定性吗？

几个月后的一天，我和科尔盖特在加利福尼亚内华达山上漫步（科尔盖特是美国最好的辐射和流体运动专家之一，是美国后期超弹工作的主要参与者，也是在利弗莫尔用计算机模拟恒星坍缩的三个物

理学家中的一个），我向他提出我问泽尔多维奇的那个问题，他也那样回答我：气流是稳定的，气体不可能生成舌头而逃脱X射线的作用。“你怎么知道呢，斯特林？”我问。“已经证明过了。”他回答。“哪儿能找到计算或者实验呢？”我问。“我不知道……”

“那太奇怪了，”我告诉斯特林，“泽尔多维奇也是这样给我说的——流是稳定的。可他也跟你一样，不告诉我证明在哪儿。”“哦，那么泽尔多维奇也知道了，真有意思。”斯特林说。

那么，现在我也知道了。我没想过要知道。但结果是不可避免的。特勒-乌拉姆思想一定是用裂变（原子弹）触发器最初产生的X射线来加热、增压，点燃超弹的聚变燃料（图6.5）。80年代，几种解了密的美国出版物证实了这确实是特勒-乌拉姆思想的一部分，要不，我也不会在这儿讲了。

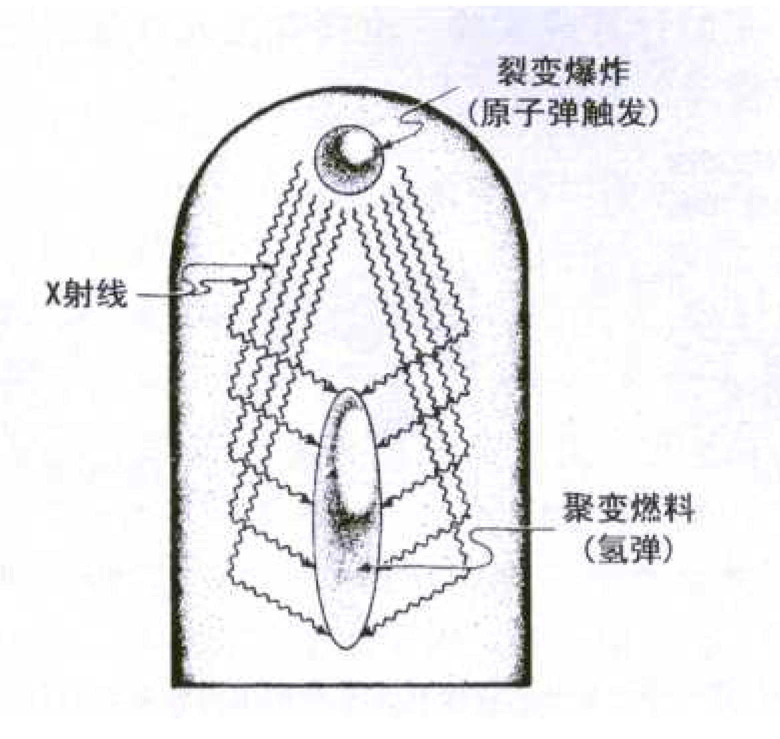


图6.5　特勒-乌拉姆和萨哈洛夫-泽尔多维奇氢弹设计思想的草图：由裂变提供能源的爆炸（原子弹触发器）产生强X射线，X射线以某种方式聚焦到聚变燃料（氘化锂，LiD），将它加热并压缩足够长的时间，产生聚合反应，X射线的聚焦技术和其他实际问题要困难得多，即使知道了特勒和乌拉姆的这点“机密”，也不过是在制造实际超弹的路上向前走了“无限小”的一步。

是什么令惠勒从黑洞的怀疑者变成了信仰者和宣扬者？恒星坍缩的计算机模拟不过是最后的激发因素。更重要的还是他想打破认识上的障碍。从20年代到50年代，这些障碍普遍存在于全世界物理学家的头脑中。它部分来自那个出现在黑洞中的史瓦西奇点，也来自奥本海默和斯尼德理想化计算的令人疑惑的结果：从外面静止的观察者看，坍缩的恒星将在临界周长处（“史瓦西奇点”）永远冻结，而从恒星表面的观察者看，坍缩会很快通过冻结点继续下去。

在莫斯科，朗道和他的伙伴们，一方面相信奥本海默和斯尼德的计算，同时却在协调这两个观点上遇到了很大的麻烦。“当人类头脑面临两个可能同时都正确的观点时，你不能想象那是多么难以理解。”[39]若干年以后，朗道最亲密的朋友栗弗席兹（Evgeny Lifshitz）这样告诉我。

1958年（也就是惠勒攻击奥本海默和斯尼德结论那年）的某一天，莫斯科收到一期刊有芬克尔斯坦（David Finkelstein）文章的《物理学评论》，[40]作者是一个不出名的博士后，来自美国一所不太有名的大学，新泽西霍伯肯的史蒂芬斯理工学院。朗道和栗弗席兹读了这篇文章。它揭示了好多东西，一切事情都豁然开朗了。1

那年，芬克尔斯坦访问了英国，在伦敦国王学院作演讲时，彭罗斯（他后来将变革我们对黑洞内部事物的认识，见第13章）曾乘火车赶来听，然后满怀热情地回剑桥。

在普林斯顿，惠勒起先有点兴趣，但没全信。等几年他才会慢慢相信。他之所以比朗道和彭罗斯迟疑，我想是因为他看得更深。他坚持认为量子引力会使核子（中子和质子）在坍缩的恒中化为辐射而脱

离坍缩，这个想法似乎不可能与芬克尔斯坦的见解协调。不过，我们以后会看到，在更深的意义上，两家观点都是对的。

那么，芬克尔斯坦的见识是什么呢？他很偶然地凭简爷的数学发现了一个描述史瓦西时空几何的新参照系。他对恒星坍缩不感兴趣，也没有把他的新参照系同星体坍缩联系起来。[42]但对别人来说，他的新参照系的含义是很清楚的，为他们带来了对星体坍缩的全新认识。

坍缩恒星外的时空几何还是史瓦西的几何，所以能用芬克尔斯坦的新参照系来描述，它与我们以前遇到过的参照系（1，2章）大不相同。那些参照系（假想实验室）都很小，而且每个参照系的各部分（顶、底、边和中心）都是相对静止的。相反，芬克尔斯坦的参照系很大，可以同时覆盖远离坍缩恒星的时空区域、恒星附近的区域以及它们之间的区域。更重要的是，这些参照系的不同部分是相对运动的；远离恒星的部分是静止的，也就是没有坍缩；而恒星附近的部分则随坍缩的恒星表面向下落。相应地，芬克尔斯坦的参照系同时可以用来描述遥远的静止观察者和随坍缩恒星一起下落的观察者所看到的恒星坍缩过程。结果，这样的描述把从远处观察到的坍缩冻结与从恒星表面观察到的持续坍缩优美地协调起来了。



大卫·芬克尔斯坦，约1958年。[Herbert S.Sonnefeld摄；芬克尔斯坦提供。]

1962年，惠勒的普林斯顿研究小组的两个成员，贝克多夫（David Beckedorf）和米斯纳（Charles Misner）构造了一组嵌入图来说明这种和谐的描述，1967年，我在为《科学美国人》写的一篇文章里将他们的嵌入图改成了下面更富想象的样子：43

从前，在一只大橡皮膜上生活着6只蚂蚁（图6.6），这些蚂蚁都很聪明，让信号球以不变的速度（他们的“光速”）在膜面上滚动来互通讯息。遗憾的是，这些蚂蚁没有计算膜的强度。

一天，5只蚂蚁刚好会聚到膜中心，他们的重量使膜往下坍缩，

自己也陷了进去，但爬得太慢，出不来了。第6只蚂蚁——他们的“宇航员”——带着望远镜在远处的安全地方。当膜塌陷时，被陷的蚂蚁向“宇航员”发了信号球，所以她能跟踪他们的命运。

膜在塌陷时发生了两件事：第一，它的表面向内收缩，把周围的事物拖向塌陷中心，就像坍缩恒星的引力将事物吸向它的中心一样。第二，膜塌下来形成一个碗，就像坍缩恒星周围的空间弯曲的形状（与图6.2对比）。

膜的表面将随着塌陷而越来越快地收缩，结果，被陷的蚂蚁等时间间隔发出的信号球要经过越来越长的时间才能被“宇航员”蚂蚁收到（这类似于从坍缩恒星发出的光的红移）。塌陷开始15秒后，发出第15号球，在这同一瞬间，陷落的蚂蚁正好被吸进临界周长。因为临界周长处的膜正好以球的运动速度（光速）收缩，所以第15号球将永远停在那儿。在到达临界周长前0.001秒，陷落的蚂蚁发出第14.999号球（只画在最后一幅图上），这个球勉强比膜的收缩快一点儿，要等到137秒后才能到达宇航员蚂蚁。过了临界周长0.001秒以后发出的15.001号球必然会被吸进高度弯曲的膜区，与那5只陷落的蚂蚁挤在一起。

但宇航员蚂蚁永远也不会知道他们挤压的情况。她永远也收不到15号和在它之后发出的信号球；而刚好在15号以前发出的球却要过很长时间才能逃出来，所以在她看来，塌陷会很慢，而且正好在临界周长处冻结。

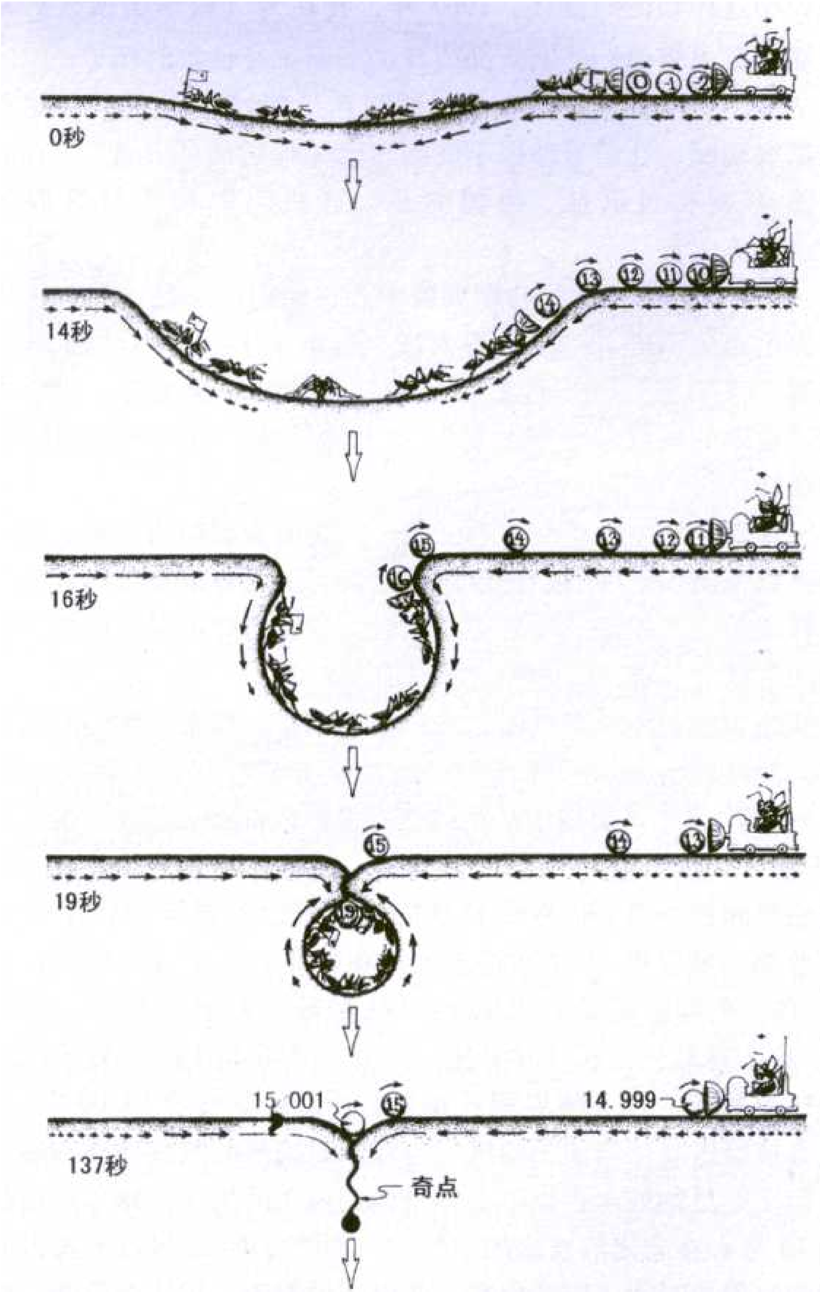



图6.6　居住着蚂蚁的橡皮膜的塌陷，为恒星形成黑洞的引力坍缩提供了一个富有想象的类比。（引自Thome（1967））

这个类比非常忠实地再现了坍缩恒星的行为：

1.膜的形状正好就是恒星周围弯曲空间嵌在嵌人图中的形状。

2.信号球在膜上的运动与光子在坍缩恒星的弯曲空间中的运动是完全一样的。特别是，在膜上的任何静止蚂蚁局部测得的信号球都以光速运动，但刚在15号之前发出的球却要过很长时间才能逃出来，在宇航员蚂蚁看来就像坍缩要冻结了。同样地，任何人局部测得的从恒星表面发出的光子也以光速运动，但刚在恒星收缩到临界周长（它的视界）前发出的光子也要过很长时间才能逃出来，在外面的观察者看来，坍缩必然显得要冻结了。

3.陷落的蚂蚁看不到任何在临界周长冻结的事情，他们会被无情地吸进临界周长，挤在一起。类似地，任何在坍缩恒星表面的人也看不到坍缩冻结，他们毫不犹豫地经历坍缩，被潮汐引力压得粉碎（第13章）。

用嵌入图来说，这就是芬克尔斯坦新参照系带来的认识。以这种方式思考坍缩就不再有什么神秘的事情了。坍缩的恒星确实会毫不犹豫地收缩到临界周长以下，远处看到的冻结表现是一种错觉。

蚂蚁比喻的嵌入图只能说明芬克尔斯坦新参照系的部分内容，更进一步的思想表现在图6.7的坍缩恒星的时空图上。

到现在为止，我们见过的惟一时空图是在狭义相对论中遇到的，例如图1.3。那张图，我们是用两种观点画的：静止在帕萨迪纳城的惯性参照系（忽略向下的引力）的观点，图1.3（c），和固定在帕萨迪纳科罗拉多林阴大道上高速奔驰的跑车上的惯性系的观点，图1.3（b）。在每幅图中，我们将选出的参照系的空间画在水平方向，时间画在垂直方向。

在图6.7中，我们选的是芬克尔斯坦的参照系，跟往常一样，这

个参照系的三维空间（“芬克尔斯坦空间”）的二维画在水平方向上，而从它观测的时间（“芬克尔斯坦时间”）画在垂直方向上。因为芬克尔斯坦参照系在远离恒星处是静止的（没有坍缩），所以在那儿的时间也就是静止观察者所经历的时间；又因为芬克尔斯坦参照系在恒星附近会随坍缩的恒星表面向下落，所以在那儿的时间也就是下落的观察者所经历的时间。

图中表现了两个水平碎片，分别示意特定时刻的两个空间维，但忽略了空间弯曲，所以它看起来是平直的。更特别的是，包围恒星中心的周长被忠实地表现在这两个水平面上，但是半径（距中心的距离）却表现不出来；为了同时忠实地表现周长和半径，我们还得用图6.2那样的嵌入图或者图6.6的蚂蚁的类比。空间曲率那时也会慢慢清楚：周长将小于2π乘以半径。将水平碎片画平，是人为抹去了它的曲率。这种错误的空间平直化是为了把图画明白所付出的代价。代价的收获是，我们能在单独的一幅图中同时看清空间和时间。

图中最初时刻（底下的水平面），缺了一个空间维的恒星是大圆的内部，假如失去的那一维还在，恒星应该是一个大球的内部。在后来时刻（第二片），恒星收缩了，成了小圆的内部。再后来，恒星通过临界周长；再后来，它就收缩到零周长，在那儿生成一个奇点，照广义相对论，恒星在那里被挤碎了，不存在了。奇点的具体情况我们要到第13章才讨论，但应该知道，它完全不同于物理学家从20年代讲到50年代的那个“史瓦西奇点”。“史瓦西奇点”是他们对临界周长和黑洞的误会的说法；这里的“奇点”是留在黑洞中心的一样东西。

黑洞本身是图中黑色的时空区域，也就是临界周长（即坍缩恒星未来的表面）以内的区域。黑洞表面（视界）就在临界周长处。

在图中还画了几个跟着恒星表面的粒子的世界线（穿过时空的轨道）。当我们从图的下方往上看时（也就是随时间流逝），会发现这些世界线越来越向恒星中心（图的中心轴线）靠拢。这个运动体现了

恒星随时间的收缩。

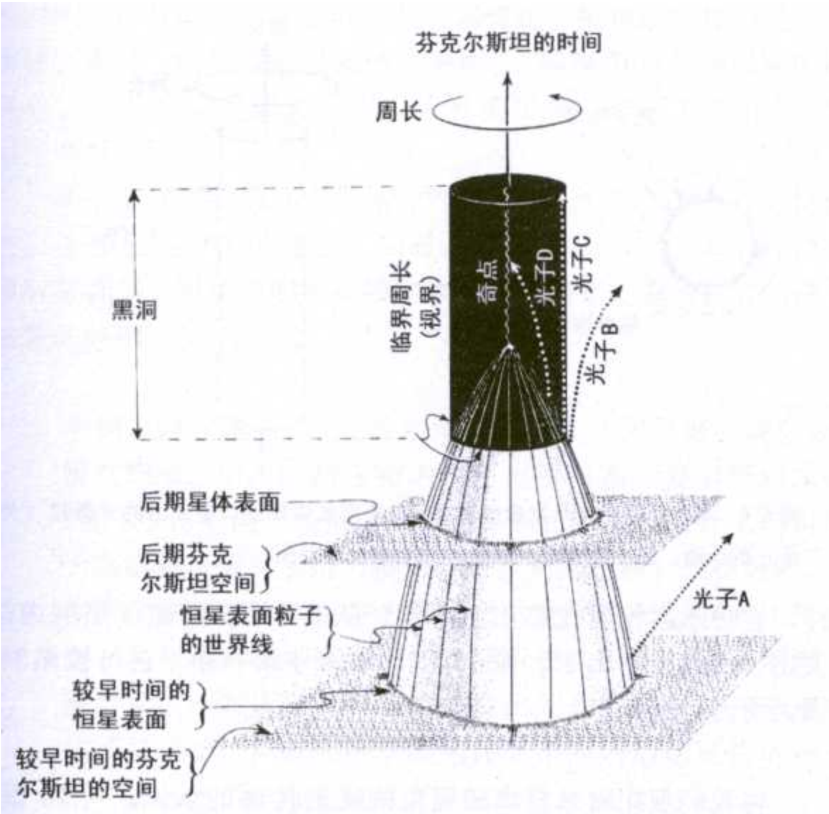



图6.7　恒星坍缩为黑洞的时空示意图。用芬克尔斯坦参照系的时间，方向向上。这个参照系的三维空间的二维画在水平方向上。水平碎片是坍缩的恒星和它生成的黑洞在芬克尔斯坦的特定时刻的二维“快照”，但空间的弯曲被抹平了。

最有意思的是四个光子（光的粒子）的世界线。这些光子类似于蚂蚁的信号球。光子A是在恒星开始坍缩的时刻从表面向外发出的（底下的空间片）。随时间流逝（从图下往上看）它很轻松地越走越远。光子B是在恒星到达临界周长前一会儿发出的，要过很长时间才能逃出去；它有点儿像蚂蚁发出的14.999号信号球。光子C是在临界周长处发出的，像15号信号球一样，它将永远停留在那儿。光子D从临界周长内部（黑洞内部）发出，永远逃不出来，像15.001号信号球

一样，它将被黑洞的强大引力拉进奇点。

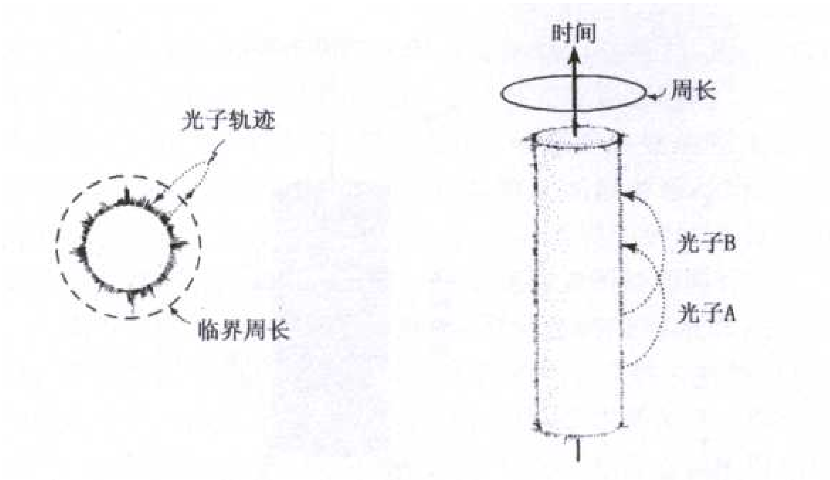



图6.8　根据牛顿物理学定律预言的从临界周长内部的恒星发出的光微粒（光子）的运动。左：空间图（同图3.1）；右：时空图。

将我们现在对来自坍缩恒星的光的传播的认识，与18世纪关于从小于临界周长的恒星发出的光的预言进行对比，是很有意思的。

回想一下（第3章），18世纪末，英国的米歇尔和法国的拉普拉斯用牛顿的引力定律和光微粒预言了黑洞的存在，那些“牛顿黑洞”实际上是周长很小（小于临界周长）的静态星体，引力阻止了星体附近的光逃逸出去。

图6.8的左边（空间图，不是时空图）描绘了这样一颗在临界周长内的恒星和从它表面近垂直（径向）发出的光子（光微粒）的空间轨道。向外飞的光子像扔出去的石块，会被恒星引力减速、停止，然后落回恒星。

右图描绘了两个光子在时空图中的运动。向上的是牛顿宇宙的时间，向外的是绝对空间。随时间流过，圆形的恒星扫过一立体柱，在任一时刻（通过图的水平面），恒星由与左图相同的圆表示。光子A

随时间向外飞，然后落回来；光子B发出稍晚，但行为是一样的。

关于临界周长以下的恒星和它发出的光子，将这个（错误的）牛顿观点与（正确的）相对论观点（图6.7）拿来对比是很有好处的。对比说明了牛顿定律和爱因斯坦定律的预言的两点显著的差别：

1.牛顿定律（图6.8）允许小于临界周长的恒星健全地存在，没有坍缩，引力的挤压被内部压力所平衡。爱因斯坦定律（图6.7）坚信，当恒星小于临界周长时，没有什么内部压力能够抵抗强大的引力挤压，恒星别无选择，只有坍缩。

2.牛顿定律预言，从恒星表面发出的光子先可以向外飞得很远，甚至跑到临界周长以外，然后被拉回来。爱因斯坦定律要求，任何从临界周长内部发出的光子总是向越来越小的周长运动。只有在恒星收缩比光子向内运动更快的情况下，这些光子才可能脱离恒星的表面（图6.7）。

尽管芬克尔斯坦的发现和原子弹的数字模拟完全令惠勒相信大质量恒星的坍缩必然产生黑洞，但坍缩的星体物质的命运在60年代仍然困扰着他，跟他1958年在布鲁塞尔遇到奥本海默时一样。广义相对论认为，恒星物质在黑洞中心的奇点处会被压碎而消失（第13章），但这个预言在物理上是不能接受的。在惠勒看来，广义相对论显然不能用在黑洞的中心，而该拿新的量子引力定律来取代它，这些新定律一定不会让恒星被压得粉碎。惠勒猜测，以他在布鲁塞尔阐释的观点为基础，也许新定律能将坍缩物质转化为辐射，通过量子力学的“隧道”而逃出黑洞散人星际空间。检验这个猜想需要更深刻地认识量子力学与广义相对论的结合，猜想的美妙也在于此。这是发现量子引力新定律的试验基础。

作为惠勒60年代初的学生，我认为他的猜想太无理了，物质在奇点转化成辐射，然后通过隧道脱离黑洞——惠勒怎么会相信这样的事

情呢？如惠勒所说，量子引力新定律在黑洞中心的奇点当然是重要的，但不是在临界周长附近。临界周长是“宏观领域”的东西，广义相对论一定是高度精确的；而广义相对论的定律也是不容争辩的——没有什么东西可以逃出临界周长。引力把握万物，所以不会有“量子力学的隧道”（不管它是什么）让辐射跑出去，我坚信这一点。

1964和1965年，惠勒和我同哈里森和若野一起写了一本关于冷死星和星体坍缩的专业书。[44]惠勒一定要在最后一章写进他关于辐射可能通过隧道逃出黑洞散入星际空间的猜想，令我很吃惊。为说服他不要在书里写这个猜想，最后我打电话向大卫·夏普（David Sharp）求助，他是惠勒的一个博士后。惠勒、大卫和我在三方电话上激烈争论，最后他终于投降了。

惠勒是对的，大卫和我错了。10年后，泽尔多维奇和霍金将用新建立起来的广义相对论与量子力学的部分结合，从数学上证明辐射能够通过隧道逃出黑洞——尽管很慢很慢（第12章）。换句话说，黑洞会蒸发，不过蒸发很慢。一个黑洞，从恒星坍缩形成到最后消失，所经历的时间将远大于我们宇宙现在的年龄。

我们为事物取的名字是很重要的。电影明星的经纪人大概很明白这一点，所以他们把诺玛·珍尼·贝克尔（Norma Jean Baker）改为玛丽莲·梦露（Marilyn Monroe），把贝拉·布拉斯柯（Béla Blasko）改为贝拉·卢戈西（Béla Lugosi）。[2]物理学家当然也明白。在电影业，一个名字定下一个基调，也就是观众评价影星的思想基础-——如梦露的魅力，卢戈西的恐怖。在物理学中，一个名词有助于建立认识物理学概念的思想参照系。好的名同会让人构想出一幅能突出概念最重要性质的认识图景，从而有助于以某种潜意识的直观的方式启发好的研究；坏名词则会产生阻碍研究的思想障碍。

在从1939年到1958年间，对物理学家们认识恒星坍缩起最大阻碍作用的，也许是他们为临界周长所取的名字：“史瓦西奇点”。“奇

点”一词会令人想象一个引力无限大的区域，我们知道的物理学定律在这儿失败了——我们现在认识了，这幅图景对在黑洞中心的物体来说是正确的，但对临界周长并不正确。由于这幅图景，物理学家难以接受奥本海默-斯尼德结论，在坍缩的恒星上通过史瓦西奇点（临界周长）的人感觉不到无限的引力，也没看到物理学定律的失败。

史西奇点（临界周长）一点儿也不奇异，芬克尔斯坦发现的新参照系让这一点完全清楚了。他的参照系说明，史瓦西奇点不过是一个位置，事物可能落到它的里面，却没有什么能从它里面跑出来——也就是说，我们在外的人不可能看到它的里面。芬克尔斯坦的参照系还表明，坍缩的恒星在沉没到史瓦西奇点以后将继续存在，就像太阳落到地平线下也仍然存在着。不过，正如我们地球上的人看不到地平线下的太阳，远离坍缩恒星的观察者在恒星坍落到史瓦西奇点后，也就看不到它了。这样的类比激发了50年代康奈尔大学的物理学家林德勒（Wolfgang Rindler），他为史瓦西奇点（临界周长）取了个新名字，一直叫到现在：他叫它视界（地平线）。

剩下的问题是，恒星坍缩生成的物体该叫什么？从1958年到1968年，东西方都用过不同的名字：苏联物理学家用的名字突出远离坍缩的天文学家的观点。回想一下，由于光为了逃脱引力的掌握需克服的巨大困难在远离坍缩的人看来是永远存在的，所以恒星表面似乎不会真的到达临界周长，视界不会真的形成。对天文学家来说（假如他们的望远镜真能看到坍缩的恒星），它似乎刚好在临界周长处冻结了。因此，苏联物理学家把坍缩生成的物体称为冻星——这个名字为他们在60年代的坍缩研究奠定了基调和认识基础。

相反，西方强调的是运动者的观点，观测者随坍缩的恒星表面下落，通过视界，到达真正的奇点，所以相应地他们将生成的物体称为坍缩星。这个名字有助于物理学家将精力集中到一点，那是后来惠勒最关心的一点：奇点的本质，量子物理学与时空曲率可能会在这里结合到一起。

两个名字都不令人满意。也没有人特别关注包围坍缩星的视界，也就是那个引起人们产生星体“冻结”的光学错觉的视界。60年代，物理学家的计算逐步揭示了视界的重要性，而惠勒——他比谁都更关心用一个优美的名字——也越来越感到不满意了。

惠勒喜欢躺在浴缸里放松或躺在床上考虑我们关于事物的名称。有时为了一个正确的名字他会这样寻找几个月。现在他要找一个词来代替“冻星”和“坍缩星”，在1967年下半年，他终于找到了一个理想的名字。

照他一贯的作风，惠勒没有去告诉他的同事们说，“我为这些东西找到了一个伟大的新名字，让我们叫它这……这个吧。”相反，他干脆自己就用了，好像本来就不存在别的名字，好像大家都同意这个名字。他公开这个命名是1967年晚秋在纽约市举行的一次脉冲星会议上，然后，12月在美国科学促进会的一个题为“我们的宇宙，已知与未知”的演讲中，他又顽强地引用了这个词。我们没有在场的人是在演讲稿上头一回碰到这个词的：“由于下落越来越快，〔坍缩恒星的表面〕将越来越快地离开〔远处的〕观察者。光向红端移动，一毫秒一毫秒地变暗，在不足1秒的时间内，就暗得看不见了……[恒星]像柴郡猫一样消失了，猫只留下它的笑，[3]而恒星只留下它的引力。引力，是的，只有引力，没有光，同样也没有出现任何粒子。而且，从外面看，光和粒子是一样的……[它们]落下黑洞只是为它补充了质量，增强了引力。”[45]

黑洞是惠勒取的新名字。几个月内，东西方的相对论物理学家、天体物理学家和普通大众就热情采纳了——只有一个例外：在法国，词组trou noir（黑洞）隐含淫秽的意思，被抵制了好多年。

[1] 实际上，包括爱丁顿在内的其他物理学家很早就在其他场合发现了芬克尔斯坦的见解，但他们没有理解它的意义，很快就把它忘了。[41]

[2] 卢戈西（1882～1956）是匈牙利出生的美国演员，他以扮演恐怖角色出名，而且

恐怖的角色也成了他生活的一部分。——译者

[3]　这儿说的是《爱丽思漫游奇境记》（Lewis Carroll著，汉译本有几种，最有趣的可能是赵元任先生1969年在美国出版的译本（商务印书馆，1988））第6章的那只柴郡（Cheshire）猫："这回它慢慢地消失，从尾巴尖儿起，一直到头上的笑脸。那笑在它全身都消失后还留了好一会儿。"（这尽管是儿童文学名著，但西方科学家也常在科学作品里引用其中的故事和语言。）——译者

# 第7章　黄金年代

黑洞被发现会自转和脉动，

它贮藏能量，

也释放能量，

而且没有毛

那是1975年，在芝加哥城南部密歇根湖畔的芝加哥大学，在一间俯瞰56号大街的办公室里，钱德拉塞卡正埋头发展黑洞的完备的数学描述。他这时分析的黑洞，与60年代初物理学家刚开始把握黑洞概念时的那些怪物有根本的不同。这中间的10年，是黑洞研究的黄金年代，是我们改变对广义相对论预言的认识的年代。

在黄金年代开始的1964年，黑洞被顾名思义地认为是空间中物体可以落进却没有东西可以出来的一个洞。但在黄金年代里，经过100多位物理学家用爱因斯坦广义相对论进行的一次又一次计算，这种图像已经改变了。现在，当钱德拉塞卡坐在芝加哥办公室里计算的时候，黑洞不再被认为仅仅是一个静悄悄的洞，而是一个活动的物体：黑洞会旋转，旋转时在它周围的弯曲时空里会产生龙卷风一样的涡旋运动。在旋涡中应该贮藏着大量的能量，自然可以开发这些能量，用它作为宇宙爆炸的动力。当恒星、行星或小黑洞落进大黑洞时，能量会使大黑洞脉动，大黑洞的视界会内外波动，像地震后地球表面上下振动一样。这样的脉动将产生引力波——在宇宙空间向外传播的时空曲率的微澜，带着黑洞和谐的乐音。

黄金年代出现的最大惊奇，也许是广义相对论坚信黑洞的所有性质都能用三个参数来预言：黑洞的质量、旋转速度和电荷。如果谁有足够的数学素养，他就能根据这三个参数进行计算。例如，他可以计算黑洞视界的形状，引力作用的强度，周围时空涡旋的细节和脉动的频率。到1975年，多数性质我们都知道了，但还不是全部。认识黑洞剩下的性质是艰苦的挑战，正好也是钱德拉塞卡喜欢的那类挑战，为了自己的追求，他在1975年接受了挑战。

在近40年的时间里，30年代与爱丁顿论战的痛苦一直郁积在钱德拉塞卡内心，他没能回到大质量恒星的黑洞命运的研究上来。在这40年里，他为现代天体物理学做出了许多基础性的贡献——关于恒星及其脉动的，关于星系的，关于星际气体云的以及关于许多其他方面的。但总的说来，大质量恒星，的命运在诱惑着他，他终于在这个黄金年代从痛苦中摆脱出来了。



1971年秋，钱德拉塞卡与研究生特奥科尔斯基（左）和莱特曼（Alan Light-man，右）在加州理工学院学生自助食堂。[Sándor J.Kovács提供。]



1975年夏，在普林斯顿大学参加会议为黑洞黄金年代“送行”的年轻人。左起，前排：Jacobus Petterson,Philip Yasskin，普雷斯，Larry Smarr,Beverly Berger,Georgia Witt,Bob Wald。二、三排：Philip Marcus,Peter DEath,Paul Schechter，特奥科尔斯基，Jim Nestor,Paul Wiita,Michael Schull,Benrard Carr,Clifford Will,Tom Chester,Bill Unruh,Steve Christensen。[特奥科尔斯基提供。]

他回到了多数是学生和博士后的研究大家庭。黄金年代是年轻人的，他们欢迎钱德拉塞卡来到他们中间，他虽已人到中年而且行为保守，却还有颗年轻的心。在长期访问加州理工学院和剑桥大学时，他常去自助食堂，周围都是轻松随便、穿戴漂亮的研究生，而他却穿着一身暗淡的灰衣服——年轻的朋友说这种颜色是“钱德拉灰”。

黄金年代是短暂的。加州理工学院的研究生普雷斯（Bill Press）曾为黄金年代命了名，而在1975年夏当钱德拉塞卡开始计算黑洞性质时，普雷斯却忙着它的“葬礼”：筹划在普林斯顿大学召开一个为期4天的只请30岁以下的研究者参加的会。[1]会上，普雷斯和他的年轻同事们一致认为，现在是转向研究其他课题的时候了。黑洞作为旋转、脉动的动态天体的图像已大概勾画出来，而理论发现的步子开始慢下来了。所有剩下的事情似乎只是补些细节。钱德拉塞卡和几个人能将它画好，而他的年轻（现在长大了）朋友们寻找新的挑战去了。钱德

拉塞卡并不满意。

## 良师：惠勒，泽尔多维奇，席艾玛

改变我们对黑洞的认识的那些年轻人是谁？是三位杰出先生的学生、博士后和学生的学生；那三位先生是：美国新泽西州普林斯顿的惠勒（John Archibald Wheeler），苏联俄罗斯莫斯科的泽尔多维奇（Yatov Borisovich Zel'dovich）和英格兰剑桥的席艾玛（Dennis Sciama）。他们通过他们的徒子徒孙，在黑洞的现代认识上留下了自己的烙印。

三个好老师，都有自己的风格。事实上，恐怕难以找到比这更鲜明的风格了。惠勒是有魅力、凭灵感的幻想家。泽尔多维奇像一个紧密团结的球队的雄心勃勃的队员和教练。席艾玛像自我牺牲的“催化剂”。在接下来的几页，我们会一个个地认识他们。

第一次与惠勒见面的情景我还记得很清楚。那是黄金年代来临两年前的1962年9月。那时，他刚相信了黑洞的概念，而22岁的我也刚从加州理工学院毕业来普林斯顿攻读博士。我很想在惠勒指导下从事相对论研究，于是惶恐地第一次敲响了他办公室的门。

惠勒教授热情微笑地招呼我，让我进了办公室，马上就开始讨论星体坍缩的奥秘（他把我看成是他尊重的同行，而不是一个十足的新手）。惠勒在谈那个年代的文章里描述了我们讨论时的心情：“在物理学史中，人们现在比过去任何时候[在星体坍缩的研究中]更有把握猜测他遇到了新现象，它有自己神秘的本性，等着人们去揭示……不论[未来研究的]结果如何，我们感到，最终会[在星体坍缩中]发现某个现象，广义相对论将在这里戏剧性地找到自己的归宿，圆满实现它与量子物理学火热的结合。”[1]一个小时后我出来时，已经完全相信了。

惠勒带着5到10名普林斯顿的学生和博士后——启发他们，但不具体指导。他认为我们很不错，能自己做好具体的事情。他为我们每个人提出了第一个研究问题——可能为星体坍缩、黑洞或广义相对论与量子物理学的“火热的结合”带来新发现的那些问题。如果发现第一个问题太难，他会慢慢将我们引向更容易的方叫；如果问题很容易，他会鼓励我们把能认识到的东西都提出来，写一篇论文，然后去迎接更有挑战性的问题。我们很快学会了同时进行几个问题——一个太难的问题在解决之前，在有希望带来巨大回报之前，会一次又一次地走进脑海，而那些简单的问题很快就会有结果。自始至终，惠勒都只是给我们提些忠告，免得错误太多，但从来不会让我们感觉是他替我们解决了问题。

我的头一个问题很怪：拿根磁铁棒，磁场穿过它，从两端出来。场由力线组成，将铁粉洒在下面有磁铁的白纸上，谁都能看到这些力线（图7.1（a））。相邻的力线相互排斥。（将两块磁铁的北极相互推近，就能感觉这种排斥。）尽管如此，每个磁场的力线还是因为磁铁而聚在一起。把磁铁拿走，排斥会使力线爆裂（图7.1（b））。这件事情，我还在读大学时就熟悉了。在普林斯顿的惠勒办公室里与他的一次私人长谈中，他又让我想起它来。他那回讲了他的朋友，塔拉哈西弗罗里达大学的梅尔文（Mael Melvin）教授的最新发现。

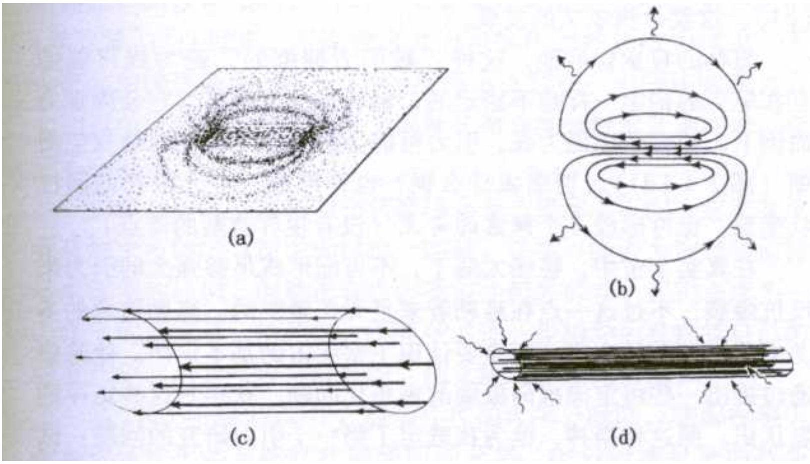



图7.1　（a）磁棒的磁力线，在磁棒上面的白纸上洒些铁粉，就能看到这些力线。（b）纸和磁棒拿走后，同样的磁力线。相邻力线间的压力使它们沿箭头方向爆裂。（c）无限长的柱状磁力线束，它的场很强，能量产生的时空弯曲（引力）足以克服力线间的排斥而将线束维系在一起。（d）惠勒猜测，如果（c）的磁力线束收缩一点，则引力足以使线束发生坍缩（波浪线）。

梅尔文发现，根据爱因斯坦场方程，磁力线不但能靠磁棒中的铁约束在一起而不会爆裂，也可以靠引力而不需要任何磁力的帮助。原因很简单：磁场有能量，而能量产生引力。（为什么能量会产生引力？回想一下，能量与质量是“等价的”（卡片5.2）：任何类型的物质（铀，氢或者别的东西）都可能转变为能量；反过来，任何类型的能量（磁能、爆发能或者别的）也能转化成物质。因此，在深层意义上，质量和能量不过是同一事物的不同名称，这意味着，由于一切形式的物质都产生引力，所以一切形式的能量也必然如此。仔细检验爱因斯坦场方程就能证明这一点。）现在，假如我们有一个极强的磁场——远比在地球上遇到的强——那么，强磁场将产生强引力，而引力会压缩磁场，将磁力线约束在一起，而不管这些力线之间有多大压力（图7.1（c））。这就是梅尔文的发现。

惠勒的直觉告诉他，这种“被引力捆绑的”磁力线可能像立在笔尖的铅笔一样是不稳定的；轻轻推一下铅笔，它会因重力而倒下。轻轻压缩磁力线，引力可能会超过压力，使力线发生坍缩（图

7.1（d））。坍缩成什么呢？也许形成一个无限长的圆柱状黑洞，也许形成一个裸露的奇点（没有视界遮蔽的奇点）。

在真实宇宙中，磁场太弱了，不可能形成足够强大的引力来反抗爆裂，不过这一点在惠勒看来是无关紧要的。惠勒追求的不是去认识存在的宇宙，而是要认识主宰宇宙的基本定律。他希望通过提出一些将定律推向极端的理想化问题，获得对这些定律的新认识。照这种精神，他为我选定了第一个引力研究的问题：试用爱因斯坦场方程来判断梅尔文的磁力线束是否会坍缩，如果会，它将坍缩成什么？

我为这个问题奋斗了好几个月。白天的战场是普林斯顿帕尔玛物理实验室的顶楼，我在那儿跟其他学物理的同学共用一间大办公室，我们一起讨论问题，互相尊重，友好和睦。晚上，回到一间小屋子，那是从第二次世界大战的士兵营房改造来的，我跟妻子琳达（画家和数学系学生）、女儿卡莱丝和我们的大柯利狗“王子”就住在这里。每天，我带着问题往返于兵营和实验楼；每过几天，就去找惠勒，请他指导。不知在纸上写了多少，在计算机上算了多少，在黑板上跟同学讨论了多少，问题终于慢慢清楚了。爱因斯坦的方程在我的穷追猛打下，终于告诉我，惠勒的猜想错了。不论用多大力量挤压梅尔文的圆柱状磁力线束，它总会反弹回来。引力永远也不可能克服磁场的排斥压力。这里没有坍缩。

这是最好的结果，惠勒热情地向我解释：当计算证实了希望，那不过是增强了一点儿对物理学定律的直觉认识。但如果计算与希望矛盾，那么我们就走上了一条通向新认识的道路。

球状星体与梅尔文的圆柱状磁力线束之间的差异是极端的，惠勒和我认识到：当球状星体非常致密时，内部引力会超过星体所能聚集的任何形式的压力。大质量球状星体的坍缩是必然的（第5章）。反过来，不论费多大力量挤压圆柱状磁力线束，不论将它的圆形截面（图7.1（d））压得多么紧密，力线束的压力总会超过引力而将力线

拉回来。圆柱状磁力线的坍缩是禁戒的，永远不会发生。

为什么球状星体和圆柱状磁场会表现得那么不同？惠勒鼓励我从每个可能的方向去探讨这个问题，答案可能会为物理学定律带来深刻的认识。但他没有告诉我该如何去探寻。我正在成为一个独立的研究者；他相信，不要他的进一步指导而发挥我自己的研究思想，对我来说是最好的。独立能培养人的力量。

从1963年到1972年，黄金年代的大部分时间，我都在努力去理解球状星体与圆柱状磁场间的差异，但只能零星地做些事情。这个问题太深太难，而另外还有些更容易的问题等着我用更多的精力去研究：星体的脉动，星体在脉动时可能发射的引力波，时空弯曲对巨大星团和对它们坍缩的影响。在这些研究中，我每年都有一两次会从抽屉里拿出几本马尼拉纸的小册子，上面记着我的磁场计算。后来，我将这些计算扩大到另一种理想化的无限长圆柱状物体的计算：热气体构成的圆柱状“星”，也就是会坍缩或者既有自转同时也坍缩的圆柱状尘埃云。尽管真实的宇宙中并没有这类东西，我零星做出的关于它们的计算却逐渐带来了结果。

到1972年，真相大白了，只有当物体在所有三个空间方向，南-北、东-西、上-下，都受到压缩时（例如，球状压缩），引力才可能超过任何形式的内部压力。相反，如果物体只在两个空间方向上经历压缩（例如，圆柱状压缩使它成为一根细长的线），引力也会增大，但还大不到战胜压力的程度。很小的一点压力，不论来自热气体还是电子简并还是磁力线，都容易超过引力，使圆柱状物体向外膨胀，如果物体只在一个方向上压缩，变成一张薄饼，压力要超过引力就更容易了。

在球形、无限长圆柱形和无限延展的薄饼的情形，我的计算是明白而确定的，对这些物体，计算容易把握。更难计算的——实际上远远超出了我的能力——是有限大小的非球形物体。但从我和我年轻同

学的计算得来的物理直觉告诉了我应该期待的东西。我提出了一个环猜想：[2]

请你随便想一类物体——颗星，一个星团，一束磁力线或别的你喜欢的东西，测量它的质量（例如，可以通过测量引力作用在轨道上行星的强度来实现），根据质量计算物体的临界周长（18.5公里乘以以太阳质量为单位的星体质量）。如果星体是球形的（实际上它不是）而且迟早会坍缩或被挤压，那么当它压缩到临界周长以下时，它会形成黑洞。假如物体不是球形的，会发生什么呢？环猜想旨在提出一个答案（图7.2）。




图7.2　照环猜想，只有当一个具有临界周长的环能够将物体圈起来时，那个坍缩的物体才能形成黑洞。

做一个周长等于那个物体的临界周长的环。然后，将物体放在环的中心，旋转环，看它是否能在各个方向包围物体，如果做到了，那么物体必然已经形成了一个包围自己的黑洞视界。如果做不到，那么物体还不够致密，成不了黑洞。

换句话说，环猜想讲的是，如果一个物体（一颗星，一个星团或者别的什么）经历高度非球形压缩，那么只有当它在各个方向的周长都小于临界周长时，它才会形成包围自己的黑洞。

我提出这个猜想是在1972年，从那时起，我一直在同别人一起努

力，想知道它是对还是错。答案藏在爱因斯坦场方程里，但事实证明想找到答案是极其困难的。同时，支持环猜想的偶然证据不断在增多。最近，1991年，康奈尔大学的夏皮罗（Stuart Sharpiro）和特奥科尔斯基在超级计算机上模拟了高度非球形恒星的坍缩，发现包围坍缩星体的黑洞的形成完全像环猜想预言的那样。假如环能在坍缩的星体上滑动和旋转，黑洞就可以形成；如果不能，就没有黑洞。但他们只模拟了几个具有特殊非球对称形状的星体，所以在我提出环猜想近四分之一世纪后，我们还是不能肯定它是否正确，不过看来它很可能是正确的。

如果说，泽尔多维奇是苏联的惠勒，那么从许多方面说，伊戈·诺维科夫（Igor Dmitrievich Novikov）就像我。1962年，当我第一次见到惠勒并在他指导下开始我的研究生涯时，诺维科夫也第一次来到泽尔多维奇身边，成为他的研究队伍的一员。

我小时候的生活简单富裕——出生和成长在犹他州洛根市一个严格的摩门教大家庭里[2]——相比之下，诺维科夫的生活则是崎岖艰难的。1937年，伊戈两岁时，在莫斯科铁道部任高级官员的父亲在斯大林的大恐怖中被陷害、逮捕，然后（没有朗道那么幸运）被处决了。母亲活下来，但被抓进了监狱，然后被流放。伊戈是一个叔叔养大的。

60年代初，我在加州理工学院学物理，伊戈是莫斯科大学的物理学研究生。

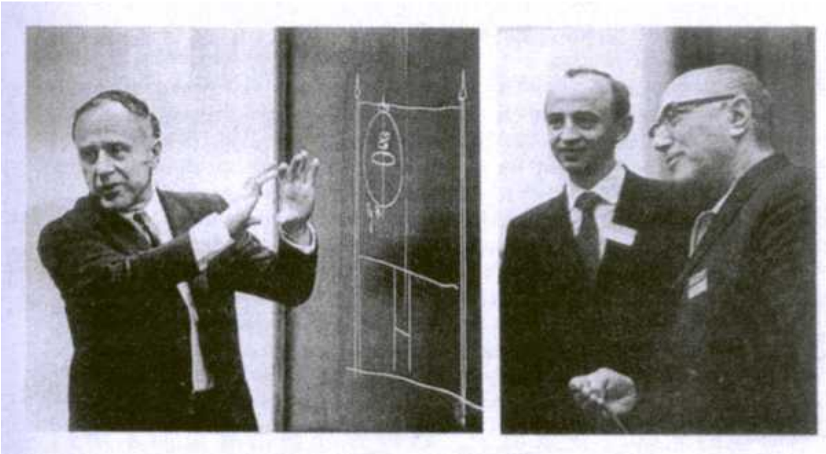
1962年，我准备去普林斯顿读研究生，跟惠勒研究广义相对论，加州的一个教授告诫我，不要学那门课。广义相对论与真实宇宙没多大关系；应该在别的地方寻找有趣的物理学挑战。（那个年代，普遍都在怀疑黑洞，没多少人对它有兴趣。）这个时候，伊戈在莫斯科以广义相对论的专题研究成了博士，他夫人诺娜也是物理学家，朋友告诫她，相对论是逆流，与真实宇宙无关。为了前途，她丈夫应该离开

它。

我对那些告诫满不在乎，还是赶到了普林斯顿。而诺娜却很担心，她利用在爱沙尼亚开物理学会议的机会，向著名物理学家泽尔多维奇征求意见。她找到泽尔多维奇，问他广义相对论是不是很重要。泽尔多维奇坚决有力地回答她，相对论对天体物理学研究来说，将变得极端重要。接着，诺娜介绍了她丈夫正在研究的一个思想：星体形成黑洞的坍缩可能类似于我们宇宙的大爆炸起源，只不过时间倒转往回流了。[3]诺娜讲着，泽尔多维奇越听越高兴，他自己有过同样的想法，而且还在探讨。

几天后，泽尔多维奇闯进了诺维科夫和许多同学在莫斯科大学史特恩堡天文研究所的办公室，一进门就向诺维科夫追问他的研究。他们尽管思想相同，但研究方法全然不同。诺维科夫已经是相对论大家了，他用了一种巧妙的数学计算来说明大爆炸与星体坍缩之间的相似性。泽尔多维奇不太了解相对论，他靠的是深刻的物理学洞察，而计算很粗。泽尔多维奇意识到，在这儿遇到跟他一样的天才了。那时候，他刚从核武器的发明者和设计者的生命中走出来，正要组建一支新的研究队伍来研究他新发现的天体物理学。作为广义相对论的大师，诺维科夫自然是理想的队员。

诺维科夫在莫斯科大学很快乐，不太想跟他走，泽尔多维奇只好施压了。他找到应用数学研究所（他的队伍要在这儿集中）所长克尔迪什（Mstislav Keldysh），克尔迪什给莫斯科大学校长彼得洛夫斯基（Ivan Petrovsky）打电话，彼得洛夫斯基请来诺维科夫。诺维科夫惶恐地走进校长的办公室。他从没想过会走进这个高居校园中心大楼的办公室。彼得洛夫斯基很坚决：“也许你现在不想离开学校去跟泽尔多维奇工作，但你会想去的。”[4]诺维科夫同意了，尽管会遇到一些困难，但他从没后悔过。



上：左，惠勒，约1970年。右，诺维科夫和泽尔多维奇，1972年。下：席艾玛，1955年。
[左上，普林斯顿大学Joseph Henry实验室藏；右上，钱德拉塞卡提供；下，席艾玛提供。]



泽尔多维奇作为年轻天体物理学家的老师，作风还是他在核武器设计小组工作时形成的："泽尔多维奇点火，队员加油"——当然，偶尔也会有别的组员闪出新的思想火花（如在相对论研究中，诺维科夫

经常如此）。然后，泽尔多维奇会热情地将年轻同事的观点拿给队员们认真辩论，让它很快成熟起来，成为他自己和发现者共享的财富。

诺维科夫生动讲述了泽尔多维奇的风格。他用名加简称的姓来称呼他的老师（在俄语中，这种称呼显得尊重而亲切），他说："雅科夫·波里希常在早晨五六点钟打电话叫醒我，'我有了一个新想法！一个新想法！快到我房间来，我们谈谈！'我去了，常常要谈很长很长的时间。雅科夫·波里希认为我们也都能像他那样长时间地工作。他和队员们常从早晨6点到10点讨论一个题目，然后换另一个题目，讨论到中午。午餐后，大家散散步，做做操，或者打个吨。喝点儿咖啡后，又更激烈地争论到五六点钟。晚上，大家自由了，可以算点儿东西，想点儿问题或者写点儿什么，为明天做好准备。"[5]

泽尔多维奇在设计核武器时养成了坏毛病，一贯要别人跟他走：照他的时间表，他工作时别人也工作，他睡觉时别人也睡觉。（1968年，惠勒、萨哈洛夫和我在苏联南端的一家旅馆的房间里跟他讨论了一个下午的物理学。激烈讨论儿个小时后，泽尔多维奇突然宣布该睡会儿觉了。然后他躺下来睡了20分钟，惠勒、萨哈洛夫和我也轻松了，各人在房间角落里静悄悄地看书，等着他起来。）

泽尔多维奇可没耐心做像我这样的完美主义者，他不在乎一定要把所有计算细节都弄对；他只关心主要概念。他能像奥本海默那样，将无关紧要的事情抛开，集中在关键问题上，而且几乎不会出错。他在黑板上画几个箭头，几条曲线，一个不到半行的方程，几句生动的点评，就把队员引到了研究问题的核心。

他能很快判断一个思想或一个物理学家的价值，却很难改变自己的判断。一个轻率的错设判断，他能坚持好多年，这常常使他看不到重要的事实。他不相信微小黑洞会蒸发，就是这样一个例子（第12章）。但如果他的迅速判断是正确的（通常是这样），它们会使他大踏步地走向知识的前沿，比我见过的任何人都更快。

泽尔多维奇与惠勒是截然不同的。泽尔多维奇以他自己源源不断的思想和与队员们共同发展起来的思想，凭着严格的管理来塑造他的队伍。惠勒为他羽翼未丰的学生提供的是一种哲学氛围，让他们感觉到处都有思想，正等着他们去采撷。但是他从不参加学生们的讨论。惠勒的最高目标是学生的教育，不在乎拖慢了他发现的步伐。而泽尔多维奇还像在超弹计划里那样，不惜多大代价也要尽可能赶在前头。

泽尔多维奇老爱在大清早可恶地打电话把别人吵醒，要人家注意他，跟他谈论，赶快往前走。惠勒在我们学生看来是世界上最忙的人，忙着他自己的计划而把我们忘了。不过，当我们需要他的指导、智慧和鼓励时，总能找到他。

那个年代的第三位良师席艾玛的风格又与众不同。60年代和70年代初，他大部分精力都用来为他的剑桥大学的学生营造一个理想的成长环境。因为他把个人研究放在第二位，所以他比那些学生还落后，总没能升为堂堂的剑桥教授（比美国的教授地位更高）。获得奖励和荣誉的更多是他的学生。到70年代末，他原来的学生霍金（Stephan Hawking）和里斯（Martin Rees）成了剑桥教授。

席艾玛是催化剂，他比学生密切接触物理学最重要的新发现。每当一个有意义的发现发表了，他都让学生去阅读，并向别人报告。每当伦敦有什么精彩的演讲，他都会带着周围的学生或让他们坐火车去听。他有很敏锐的感觉，知道哪些思想有意义，什么问题值得探索；启动研究计划该读什么书，寻求技术指导该找哪些人。

一种忘我的渴望驱动着席艾玛去认识宇宙是如何构造的。他自己说这个驱动是一种形而上学的焦虑。宇宙似乎太疯狂，太怪异，也太迷人，接近它的惟一途径是想办法认识它，而认识它的最佳路线是依靠他的学生。让学生解决最艰难的问题，他可以比自己停留在这些问题上更快地走向下一个问题。

[1] 据比尔·普雷斯的同胞特奥科尔斯基（Saul Teukolsky）的回忆，“这个会是比尔对他所谓的受了刺激的反应。那时在开另一个会，我们都没被邀请，但所有灰色大人物都参加了，于是比尔决定为年轻人开一个会。”

[2] 80年代末，根据母亲的建议，全家要求脱离摩门教，以反对教会对女权的压制。

## 黑洞没有毛

黄金年代的一大发现是“黑洞无毛”。科学中，有些发现是个人很快完成的，有些发现较慢，是许多研究者多方面贡献的结果。黑洞无毛属于第二种。它是三位大师泽尔多维奇、惠勒和席艾玛的学生们和其他许多人研究的产物。在接下来的几页里，我们会看到，无数研究者一步步地建立起黑洞无毛的概念，然后证明它、把握它的意义。

“黑洞无毛”的第一个线索1964年来自金兹堡，他曾为苏联氢弹发明了LiD（氘化锂）燃料，因为妻子有参与谋害斯大林的嫌疑，他没能进一步参加氢弹设计工作（第6章）。加州理工学院的天文学家刚发现类星体，宇宙最遥远的一种奇异的爆发天体时，金兹堡正在研究类星体的能源从哪儿来（第9章）。他想，一种可能是来自磁化的超大质量恒量形成黑洞时的坍缩。这种恒星的磁力线具有图7.3（a）上面所示的形状，也就是地球磁力线的形状。金兹堡猜想，恒星坍缩时，磁力线会被强烈压缩，然后猛烈爆炸，放出巨大能量。这可能有助于解释类星体。

通过计算星体坍缩的全部细节来检验这个猜想，是极端困难的，因此金兹堡走了另一条路，完成了他的第二大发现。他像奥本海默第一次粗略计算恒星坍缩所发生的事情一样，检验了一个静态恒星序列，序列中的恒星一个比一个紧密，而且穿过内部的磁力线数目都一样多。金兹堡认为，这个静态恒星序列应该能够模拟一颗单独的恒星。他导出了描述序列中每颗恒星的磁力线形状的公式，发现了令他大吃一惊的事情。当恒星接近临界周长开始形成包围自己的黑洞时，引力将把磁力线吸到恒星表面，使它紧紧贴在表面上。黑洞形成后，被吸附的磁力线就都落进视界了，没有留下能伸到洞外的（图7.3（a））。这对金兹堡的类星体动力思想并不是好消息，但确实提出了一种有趣的可能：磁化星体坍缩成黑洞时，黑洞仍然可能像根本

没有磁场那样诞生。6



左：发现第一个"无毛猜想"证据的金兹堡（约1962年）。右：第一个严格证明"无毛猜想"的伊斯雷尔（1964年）。[左，金兹堡提供；右，伊斯雷尔提供。]

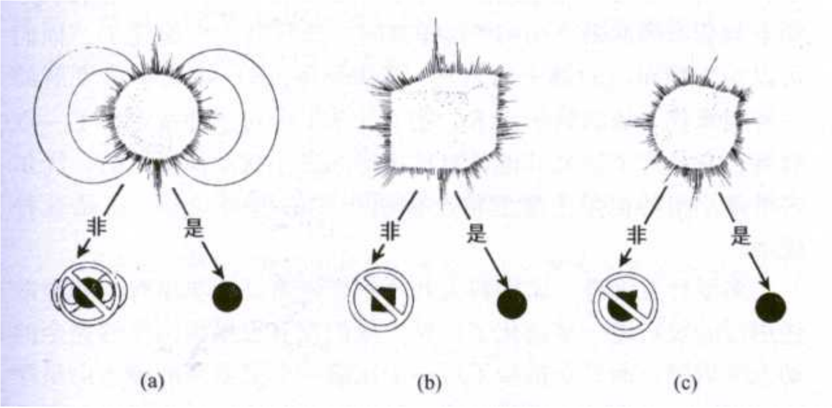



图7.3 "无毛猜想"的几个例子：（a）磁化恒星坍缩形成的黑洞没有磁场。（b）方形恒星坍缩形成的黑洞是圆的，而不是方的。（c）表面有山脉的恒星坍缩形成的黑洞没有山脉。

大约在金兹堡发现的同时，在几公里外的莫斯科，泽尔多维奇的

研究小组——诺维科夫和多罗什克维奇（Andrei　Doroshkev-ich）领头——开始问自己："圆恒星坍缩形成圆黑洞，那么变形的恒星会产生变形的黑洞吗？"举一个极端例子，方形星产生方形洞吗？（图7.3（b））。计算假想的方形星的坍缩是很困难的，于是多罗什克维奇、诺维科夫和泽尔多维奇找了一个更容易的例子：表面上隆起一座小山的近球状恒星在坍缩后形成的黑洞视界面上也会有隆起的小山吗？这样提问，计算就大大简化了，他们可以用惠勒和他的博士后雷吉（Tullio　Regge）在几年前发展的一种叫微扰方法的数学技术。在卡片7.1中对这种技术作了一点解释，它是为了研究其他球对称情形的微小扰动而提出的。泽尔多维奇小组的恒星上隆起的小山所产生的引力变形，正是这种扰动。

多罗什克维奇、诺维科夫和泽尔多维奇还用奥本海默和金兹堡用过的技巧进一步简化了计算：他们没有去模拟山星的完全的动力学對缩，而只是检验了后一个比前一个更紧密的静态山星序列。利用这个技巧和微扰方法，经过激烈争论，他们很快发现了一个显著的结果：当静态山星小到足以形成包围自己的黑洞时，黑洞视界必然是圆的，没有隆起的小山（图7.3（c））。[7]

类似的一个诱人的猜想是，如果坍缩的方形恒星要形成黑洞，它的视界也会是圆的而不会有棱角（图7.3（b））。假如猜想正确，那么黑洞就不会留下什么证据来证明产生它的恒星是方的、圆的或者是有山的，而且，（照金兹堡的看法），也不会有证据证明原来的恒星是磁化的还是无磁性的。

7年以后，当猜想逐渐显得正确时，惠勒发明了一个精炼的短语来描述它：黑洞无毛——"毛"说的是那些可能探出黑洞泄露形成它的原来恒星的细节的东西。

惠勒的大多数同事都难以相信，这位保守高贵的先生会知道他的短语的淫邪意思。但我怀疑他知道；我在私下里偶尔也见过他玩儿恶

作剧。[1]惠勒的词儿很快传开了，不过还有来自帕斯特纳克（Simon Pasternak）的反对。他是《物理学评论》主编，多数西方黑洞研究都发表在这家杂志。大约1971年，当伊斯雷尔想在论文中用这个词时，帕斯特纳克回信告诉他，不论什么条件下，他都决不允许这种猥亵的字眼儿出现在他的杂志上。但帕斯特纳克没能挡住滚滚而来的“无毛”论文。在法国和苏联，惠勒那个词的法语和俄语译名也是令人厌恶的，遭到了更长时间的抵制。然而，到70年代后期，这个词已经在全世界各种语言的物理学家中公开使用了，没人笑话。

卡片7.1　为喜欢代数的读者解释微扰方法

在代数学里，我们学过用下面的公式计算两个数a和b的和的平方：

$$(a+b)^2=a^2+2ab+b^2$$

假定a是一个大数，例如1 000，而相比之下b很小，例如是3，那么公式中的第3项$b^2$比起其余两项来是非常小的，因而可以舍去而不会产生多大误差：

$$(1\,000+3)^2=1\,000^2+2\times1\,000\times3+3^2=1006\,009$$

$$\approx1\,000^2+2\times1\,000\times3=1\,006\,000$$

微扰方法就以这种近似为基础。$a=1\,000$就相当于一个完全球状的恒星，$b=3$像恒星上的小山，而$(a+b)^2$就是恒星和小山共同产生的时空曲率。计算曲率时，微扰法只保留山的特性的线性效应（如效应$2ab=6\,000$，在$b=3$时是线性的）；这些方法抛弃了山的所有其他影响（如$b^2=9$的影响）。只要山比恒星小得多，微扰法是高度精确的。不过，假如山长得与恒星一样大（为了使恒星变成方的，就需要这样的山），那么微扰法将产生严重的误差——就像在上面的公式里$a=1\,000$ 而$b=1\,000$：

$$(1\ 000+1\ 000)^2=1\ 000^2+2\times1\ 000\times1\ 000+1\ 000^2$$

$$=4\ 000\ 000$$

$$\neq1\ 000^2+2\times1\ 000\times1\ 000=3\ 000\ 000$$

两个结果大不相同。

金兹堡、多罗什克维奇、诺维科夫和泽尔多维奇提出无毛猜想，并为它收集证据，是在1964～1965年的冬天。广义相对论专家每3年都要在世界的某个地方团聚一次，开一个星期的会，交流思想和研究结果。第4次这样的会6月在伦敦举行。

泽尔多维奇的队员从没有走出社会主义国家的圈子，他自己当然更不行了，他刚离开核武器研究没多久。不过，诺维科夫还年轻，没进氢弹计划，他的广义相对论修养在队里是最好的（这也是泽尔多维奇招他进来的主要原因），现在成了队长（泽尔多维奇是教练），而且，他的英语虽远未精通，但也还过得去。会当然应该选他去参加。

这是东西关系良好的时期。自斯大林12年前去世以来，苏联科学家与西方同行之间的通讯往来和相互访问逐渐在恢复。自然，现在每个重要的国际会议，苏联都会派一个小规模的科学家代表团参加，它的重要不仅在于保持苏联的科学力量，也是为了向西方科学家证明苏联的力量。自沙皇时代起，俄罗斯的官僚们在西方人面前就很自悲，现在他们可以在西方公众面前因为祖国的成就而骄傲地昂起头来。这对他们来说才是更重要的。

这样，泽尔多维奇很容易地说服了当权者让他年轻的同事诺维科夫参加苏联代表团，他已经从伦敦为他争到了一个在相对论会议上作大会报告的邀请。诺维科夫有许多激动人心的事情要报告，他将深刻地体现苏联物理学的力量。

在伦敦，诺维科夫向全世界的300名杰出的相对论物理学家作了一个小时的演说，精彩极了。关于山星引力坍缩的结果只是演说的一部分，其余部分对我们认识相对论引力、中子星、星体坍缩、黑洞、类星体本质、引力辐射以及宇宙起源，是同样重要的贡献。我坐在下面听着，被泽尔多维奇小组研究的广泛和力量惊呆了。我从没见过这样的事情。[8]

诺维科夫讲完后，我同大家热烈地围着他，我发现我的俄语比他的英语还好一点儿，太令我高兴了，我成了讨论的翻译。人散后，诺维科夫和我一起出来，接着讨论。这样，我又多了一个好朋友。

我和大家都不可能在伦敦完全消化泽尔多维奇小组关于“黑洞无毛”的分析。细节太复杂了，我们在等着有人来为这项研究写篇文章，把那些细节认真地写出来。

1965年，用俄文写的东西传到了普林斯顿。[9]我又庆幸自己曾在大学时费了好多讨厌的时间学俄语。文章分析包括两部分。第一部分，显然是多罗什克维奇和诺维科夫的工作，从数学上证明了有小山隆起的静态恒星越来越紧密时，只能产生两种结果。要么生成一个包围自己的黑洞，要么产生巨大的时空曲率。由于在恒星接近临界周长时山的影响不再是“小扰动”了，这时微扰的计算方法无效，坍缩的结果也就无从知道了。第二部分我很快知道是“泽尔多维奇式”的论证：如果山原先很小，那么从直觉看，它显然不可能在恒星接近临界周长时产生巨大的时空曲率。我们必须抛弃这种可能性。另一种可能性一定是真的：恒星必然生成一个完全球形的黑洞。

在泽尔多维奇从觉看来是显然的事情（最终也将证明是对的），对多数西方物理学家来说则远不是显然的。于是，争论卷进来了。

一个有争议的研究结果有很大的吸引力，就像野炊吸引蚂蚁那样。于是，泽尔多维奇小组的无毛证据把物理学家吸引过来了，先是

一个一个的，然后是一群一群的。

第一个是伊斯雷尔，他生在柏林，长在南非，在爱尔兰学相对论，现在想在加拿大埃德蒙特发起一个相对论研究小组。伊斯雷尔用绝妙的数学技巧改进了那篇苏联论文的第一部分，即多罗什克维奇和诺维科夫研究的部分：他不仅像苏联人那样考虑了小山的情况，还考虑了任意形状和大小的山。事实上，他的计算对任何坍缩都是正确的，即使恒星是非球形的，甚至是方形的，而且计算还允许坍缩是动态的，而不仅是理想化的静态的恒星序列。伊斯雷尔的结论同样令人瞩目，它类似于多罗什克维奇和诺维科夫的结论，但要强得多："一个高度非球形的坍缩只能有两个结果：要么根本不产生黑洞，要么产生完全球形的黑洞。"不过，要让这个结论正确，坍缩体必须具有两个特殊性质：它必须不带一点电荷，必须一点也不旋转。在下面我们会明白这是为什么。[10]

伊斯雷尔第一次提出他的分析和结果是1967年2月8日在伦敦国王学院的一个演讲中。演讲很难理解，但剑桥的席艾玛却急着让学生到伦敦来听。他的学生埃利斯（George Ellis）后来回忆说，"那是一个非常非常有趣的演讲。伊斯雷尔证明了一个完全在无意中出现的定理，谁也没有想到过它；谁也没有做过这样的事情。"在伊斯雷尔快讲完的时候，米斯纳（曾是惠勒的学生）站起来提出一个猜想：如果坍缩的恒星要旋转、有电荷，会发生什么事情呢？大概也只有两种可能：或者根本没有黑洞，或者生成一个完全由坍缩星体的质量、自旋和电荷决定的具有惟一形式的黑洞。最终证明这个答案是对的，不过要等到泽尔多维奇的直觉经过检验以后。

你大概记得，泽尔多维奇、多罗什克维奇和诺维科夫曾研究过变形不大的恒星，也就是有小山隆起的近球状恒星。他们的分析和泽尔多维奇的论断引出了太多的问题：

如果坍缩恒星的表面有小山，那么坍缩的结果是什么？是山在恒

星接近临界周长时产生巨大的时空曲率呢（泽尔多维奇凭直觉否决了这个结果）？还是山的影响消失而留下一个完全球形的黑洞（泽尔多维奇喜欢这个结果）？另外，如果完全球形的黑洞形成了，它用什么办法使自己摆脱山的引力影响呢？什么使黑洞成为球形的？

作为惠勒的学生，我考虑过这些问题。不过，我没有把它们作为对我的挑战，而是让我自己的学生来迎接这个挑战。那是在1968年，我在普林斯顿读完了博士又回到了母校加州理工学院，先是博士后，现在是教授了；我也学着惠勒在普林斯顿那样开始在我周围组织自己的学生队伍。

来自布鲁克林的普赖斯（Richard　Price）是一个200磅重的健壮小伙子，留着粗胡须，系着黑色的空手道绸带。他跟我做过几个小项目的研究，包括运用微扰论的数学方法解决那些问题。现在他成熟了，能够承担更具挑战性的项目。泽尔多维奇的直觉的检验似乎是很理想的，但那只是一种情况。那是个热点，别的人在别的地方正在跟它斗，蚂蚁正一群群地向着野炊进攻。普赖斯得赶快去。

可惜他来晚了，别人赶到前头去了。他是第三个得到答案的，跟在诺维科夫和伊斯雷尔后面，[11]但他来得更坚决，更彻底，也更有远见。

普赖斯的远见经史密斯（Jack　Smith）的笔而不朽了。史密斯是《洛杉矶时报》一个幽默的专栏作家，在1970年8月27日的《时报》上，他讲述了前一天访问加州理工学院的情景："在教工俱乐部吃过午饭后，我独自在校园里转。我能感觉飘在空气中的深刻思想，即使在夏天它也在吹动着橄榄树枝。我向一扇窗户看去，黑板上盖满了方程，像路上厚厚的落叶。还有三个英文句子：普赖斯的定理：能被辐射者被辐射。舒茨的观察：被辐射者能被辐射，是能被辐射者当且仅当是被辐射者。我走过窗户，好奇地想今年秋天当一年级女大学生第一次走进校园时，它会给加州理工带来什么样的影响？我想，他们是

不会有一点儿害处的罢……我隐约感到，他们总会辐射出光芒的。”

这段话需要一点解释。“舒茨的观察”是玩笑，而普赖斯的定理，“能被辐射者被辐射”，是彭罗斯1969年一个猜想的严格证明。

我们用山星的坍缩来说明普赖斯的定理。图7.4示意了这种坍缩。左图是第6章图6.7引进的那类时空图；右图是恒星和视界的形状在不同时刻的快照序列，最早时刻在下，最晚时刻在上。

恒星坍缩时（图7.4的下面两幅快照），山越长越大，对恒星的时空曲率产生越来越强的山形扰动。然后，当恒星进入临界周长内，生成包围自己的黑洞视界时（中间那幅快照），扰动的时空曲率使视界发生形变，产生一个山形隆起。不过，这个视界的隆起坚持不了多久。产生它的星体的山这时已经沉没在黑洞以内了，所以视界不会再受它的影响，也就不再有来自山的力量使它能保持隆起。视界只能靠一种方法让隆起消失：将隆起转化为在所有方向向外传播的时空曲率的波澜（引力波——见第10章）（上面两幅快照）。有些波会落进黑洞，其余的会飞向周围的宇宙，而它们离开后，留下的就是一个完全球形的黑洞。

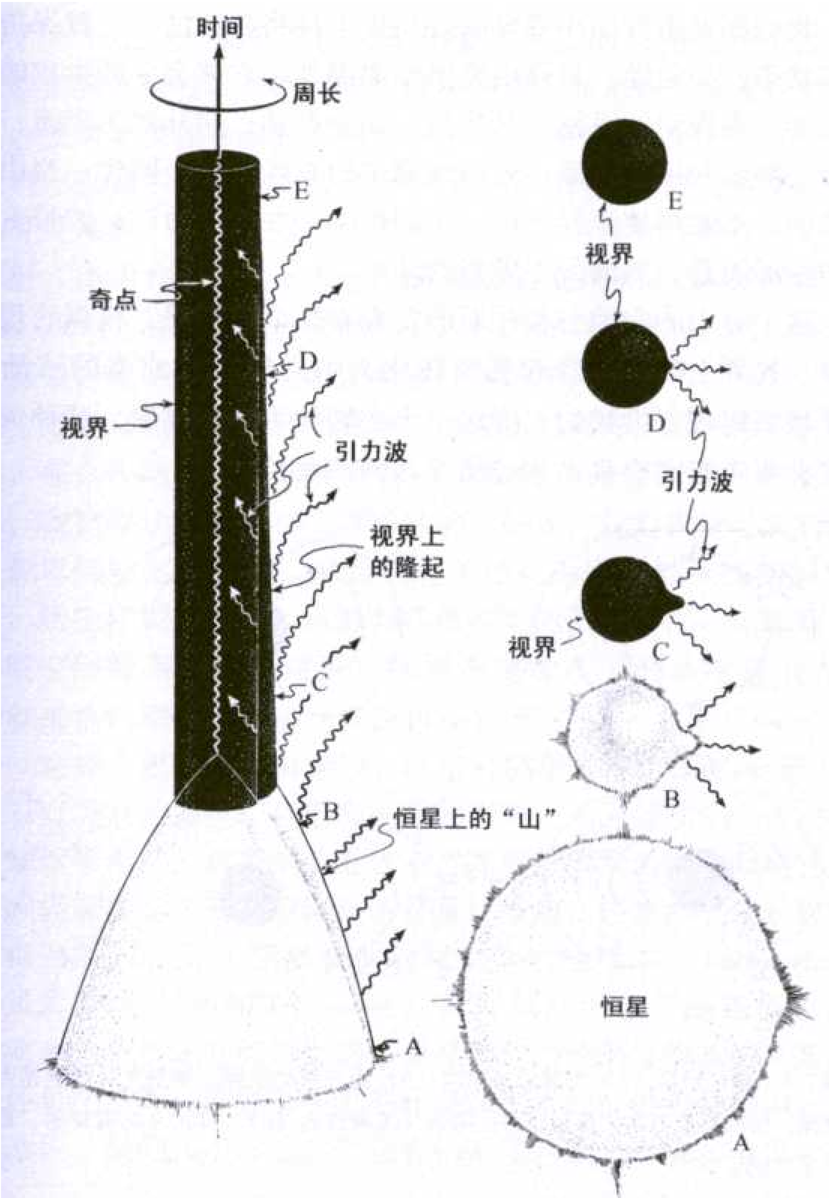



图7.4　时空图（左）和快照序列（右），说明了有山隆起的恒星形成黑洞的坍缩过程。

我们都熟悉拨动小提琴弦。只要手指将弦拉住，它就保持变形的状态；类似地，只要山突出在黑洞外，它就会令新生成的视界变形。当我们的手指离开琴弦，弦会振动，向外发出声波；声波带走琴弦变

形的能量，弦回到原来的直线状态。同样，当山沉入黑洞，不能再使视界变形，黑洞振动，发出引力波；波带走视界的变形能量，黑洞处于理想的球形。

这个有山的坍缩与普赖斯的定理有什么关系呢？根据物理学定律，视界上的山形隆起能够转化为引力辐射（曲率的波动）。而普赖斯定理告诉我们，隆起一定会转化为引力辐射，这种辐射一定会将隆起完全带走。这就是黑洞无毛的机制。

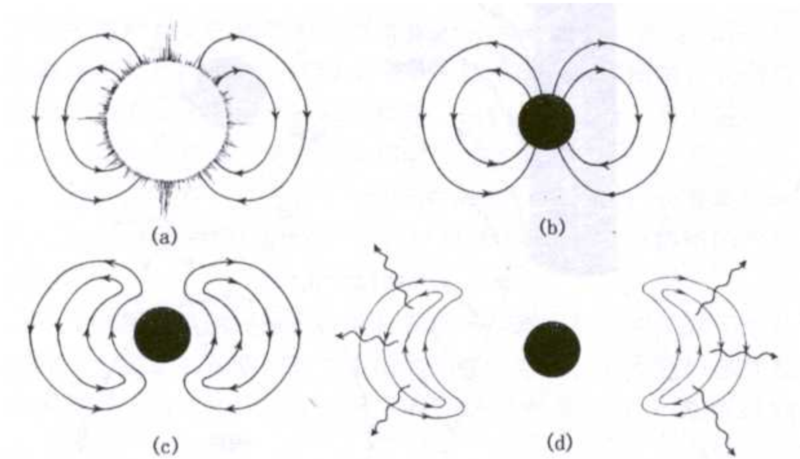



图7.5　磁化恒星（a）坍缩形成黑洞（b）的一系列快照。黑洞先从恒星得到磁场，但没有能力将它保留下来。磁场从黑洞逃走（c），转化为电磁辐射，散开了（d）。

普赖斯定理不仅告诉我们变形的黑洞如何失去变形，还告诉我们磁化的黑洞如何失去磁场（图7.5）。（这种情况的机制在普赖斯定理以前就由伊斯雷尔和他的两个加拿大学生德拉克鲁兹（Vicente de la Cruz）和切斯（Ted Chase）通过计算机模拟弄清楚了。[12]）磁化黑洞是磁化恒星坍缩后生成的。在视界吞食坍缩的恒星前，磁场被严格约束在恒星内部，那儿的电流使它不能逃逸。恒星被视界吞没后（图7.5（b）），磁场不能再受恒星电流影响，也就不再受它的约束。现在，它不仅可以穿过恒星，还可以穿过视界，视界是留不住它的。物理学定律允许场转化为电磁辐射（电力和磁力的波动），而普

赖斯定理正好要求它那么做（图7.5（c））。电磁辐射散开了，一部分落回黑洞，其余部分离它远去，留下一个没有磁化的洞（图7.5（d））。[13]

如果真像我们看到的那样，山因辐射消失，磁场因辐射而散尽，那么还能留下什么呢？什么不能转化为辐射呢？答案很简单：在物理学定律中有一类特殊的守恒律，它们肯定了某些量不可能以辐射方式振动，因此不可能转化为辐射而从黑洞附近消失。这些守恒的量是，由黑洞质量产生的引力作用，黑洞旋转产生的空间旋涡（下面讨论）和黑洞电荷产生的辐射状电力线，也就是指向洞外的电场（下面讨论）。[2]

这样，根据普赖斯定理，在所有辐射散尽以后还能留下的，就只有黑洞的质量、自旋和电荷的影响。黑洞的其他一切特征都被辐射带走了。这意味着谁也不能靠测量黑洞最后的那些性质来揭示坍缩成这个黑洞的恒星的特征，当然，恒星的质量、自旋和电荷例外。甚至（根据惠勒的两个学生哈特尔（James Hartle）和贝肯斯坦（Jacob Bekenstein）的计算）谁也不能根据黑洞的性质来判别形成它的恒星的构成是物质的还是反物质的，是质子的还是电子的，或者是中子的还是反中子的。借惠勒的话，更准确地说，黑洞几乎无毛，它仅有的毛是质量、自旋和电荷。

最后严格证明黑洞无毛（除了质量、自旋和电荷）的并不是普赖斯。普赖斯的分析严格限制在非常接近于球形而且旋转（即使有的话）非常缓慢的恒星的坍缩，他用的微扰法需要这些限制。为认识高度变形、快速旋转的坊缩恒星，需要完全不同于微扰方法的数学工具。

席艾玛在剑桥的学生掌握着需要的工具，但是太难了，难极了。为了用这些工具完全严格地证明黑洞无毛——即使在黑洞快速旋转而且因旋转而强烈变形的情况下，它的最终性质（在所有辐射散尽以

后）由质量、自旋和电荷惟一确定，席艾玛的学生们用了15年的时间。证明的大部分成绩归功于席艾玛的两个学生，卡特尔（Brandon Carter）和霍金，也归功于伊斯雷尔；但罗宾逊（David Robinson）、邦庭（Gray Bunting）和玛泽尔（Pavel Mazur）也有过重要贡献。[14]

在第3章，我评说过我们真实宇宙的物理学定律与怀特的史诗性小说《过去和未来的国王》里蚂蚁社会的物理学定律之间的巨大差别。怀特的蚂蚁信奉“凡不被禁止的事情都是必然的”，但物理学定律大大地违背了这句箴言。许多物理学定律允许的事情实际上可能性很小而从来没有发生过。普赖斯定理则是一个显著的例外，在物理学中我还很少遇到这种情形，蚂蚁的箴言实现了：假如物理学定律并不禁止黑洞以辐射排出些东西，那么辐射也就是必然的了。

黑洞最后的“无毛”状态，同样具有异乎寻常的意义。正常情况下，物理学家为了认识我们周围复杂的宇宙，常建立一些简化的理论或计算模型。为认识天气，大气物理学家建立了地球大气环流的计算模型；为认识地震，地球物理学家建立了滑动岩体的简单理论模型；为认识恒星坍缩，奥本海默和斯尼德在1939年建立了一个简化的理论模型：一团完全均匀的、没有压力的球形坍缩的物质云。我们在建立这些模型时，完全知道它们的局限，它们不过是“真实”宇宙中“在那儿”表现出的大量复杂性的一些苍白的图画。

对黑洞来说——或者，至少当辐射散尽，带走了所有的“毛”时，情况就不同了。这时，黑洞简单极了，我们可以用简单精确的数学公式来描述它，根本用不着什么理想化。在宏观世界（也就是在大于亚原子粒子的尺度上）的任何其他地方，都不会是这样的。在其他任何地方我们的数学也不会如此精确；在其他任何地方，我们也不可能从理想化模型的局限中解脱出来。

为什么黑洞与宏观宇宙中的其他物体有那么大的不同？为什么它

们（而且只有它们）会如此简单？假如我知道答案，它可能会告诉我关于物理学本质的一些更深刻的东西。可是我不知道。也许下一代物理学家会找到它。

[1] 在公开场合，我只见过一次。1971年，惠勒70岁时，正好出席在哥廷根一个城堡的高级宴会——宴会是为一个国际会议举办的，不是为他的生日。为庆祝生日，惠勒在他的椅子后挂了串鞭炮，引得邻座大乱。

[2] 80年代末，我们发现量子力学定律能引出另外的与“量子场”（将在12章讨论的一类场）相联系的守恒量；由于像黑洞的质量、自旋和电荷这样的量不能被辐射，它们将在黑洞诞生时作为“量子毛”而保留下来。虽然量子毛可能会强烈影响一个在蒸发的微观黑洞（第12章）的最终命运，但它不会对本章和以下几章的宏观黑洞（比太阳重的黑洞）产生什么影响，因为量子力学在宏观尺度上一般是不重要的。

## 黑洞旋转与脉动

无毛黑洞的性质是什么？广义相对论的数学对哪些性质能有那么好的描述？

如果黑洞是理想化的，绝对没有电荷和自旋，那么它正好就是我们在前些章遇到的球形黑洞，在数学上，由史瓦西1916年的爱因斯坦场方程的解来描述（第3章、第6章）。

如果电荷落进黑洞，那么黑洞获得一个新的特征：像刺猬一样辐射状向外竖立的电力线。电荷如果是正的，这些力线将从洞中把质子挤出来而吸引电子；如果是负的，力线会排斥电子而吸引质子。这类带电黑洞在数学上精确地由德国物理学家雷斯纳（Hans Reisner）和荷兰物理学家诺德斯特勒姆（Gtmnar Nor-dStrφm）分别在1916和1918年得到爱因斯坦场方程的解描述。但是，以前没人能理解雷斯纳和诺德斯特勒姆解的物理学意义。到1960年，惠勒的两个学生，格雷弗斯（John Graves）和布雷尔（Pieter Brill）才发现它描述了带电的黑洞。[15]

卡片7.2　苏联和西方的科学组织：对比与结果

在我和我年轻的物理学同事们努力发展环猜想，证明黑洞无毛和发现它们如何失去毛时，我们也看到，物理学的组织在前苏联与在英国和美国有多么大的不同，这些差别又产生了多么深远的影响。我们的认识可能有益于某些未来计划，特别是在前苏联，那儿的一切国家机构——不仅政府的和经济的，也包括科学的——现在（1993年）都正照着西方的路线重新组织。西方模式并不是完美的，而前苏联的体制也不一味是坏的！

在美国和英国，不断有能干的年轻人从惠勒或席艾玛那样的研究

小组中走出来。本科生可以在最后一年加入这类小组，但以后还要去读研究生。研究生可以参加3～5年，然后到另一处去读博士后。博士后两三年后要独自到另一个地方去组织自己的研究小组（如我在加州理工学院），或者加入别的研究小组。在英国和美国，不论谁有多能干，他几乎都不可能老是呆在导师身边。

在苏联情形就不同了。突出的年轻物理学家（像诺维科夫）通常跟在老师身边一二十年，有的甚至三四十年。像泽尔多维奇和朗道那样的苏联大家部在科学院的研究所工作，不在大学里，教学负担很轻，有的还没有课。为了留下以前的优秀学生，他们会组织一个永久性的研究队伍，这样的队伍团结紧密，力量强大，有的可能会坚持到导师科学生涯的终点。

我的一些苏联朋友将这种差异归结为英美体制的失败：几乎所有英美大物理学家部在大学，那里通常是研究促进教学，而要建立一支强大持久的研究队伍却缺乏足够的固定职位。结果，在英国或美国都没有产生过一个像30年代到50年代的朗道小组和60年代到70年代的泽尔多维奇小组那样的理论物理学研究群体。在这个意义上，西方是没有希望同前苏联竞争的。

我的一些美国朋友说这种差异是因为前苏联体制失败了：在前苏联，从逻辑上讲很难从一个研究所迁到另一个研究所，从一座城市迁到另一座城市，所以年轻的物理学家们只好留在老师身边；他们没有机会出去创立自己的独立研究小组。结果，批评者认为这是一种封建体制，老师像主人，而队员是他的奴隶，大部分生涯都卖给了他。主人与奴隶以复杂的方式相互依存，但谁是主人却容不得商量。如果主人是泽尔多维奇或朗道那样的大师，那么这个主人和奴隶的团队可能会产生丰硕成果。如果主人是一个独裁者而又没多大本事（通常都是这种情况），结果就惨了：糟蹋了人才。

在前苏联体制下，每一个像泽尔多维奇那样的大师都只有一个研

究队伍，尽管强有力，但还是不能与西方的相比。不同的是，在美国或英国，像惠勒和席艾码那样的大师会组织许多较弱小的研究队伍，分散在不同的领域，而最终都将对物理学产生重大影响。英美的导师们有源源不断的年轻人，帮助他们保持活跃的思想和头脑。苏联导师们难得重新开始，那时他们只好斩断与过去队伍的联系，可能会造成巨大的伤害。

实际上，泽尔多维奇就经历过这种事情：1961年，他开始建立他的天体物理学小组；到1964年，它成为全世界最优秀的理论天体物理学小组；后来，在1978年，黄金年代刚刚过去，这个小组就痛苦地分裂了，几乎每个组员都各奔前程，而泽尔多维奇要另走一条路，虽然在心理上也受到伤害，但他摆脱了障碍，可以自由地另辟新天地了。悲哀的是，他的重建不会成功。他不可能再有像从前他在诺维科夫协助下所带领的那么一群能干有力的队员团结在身边了。但诺维科夫现在独立了，在80年代，他成了一个重建的研究队伍的天才领导者。

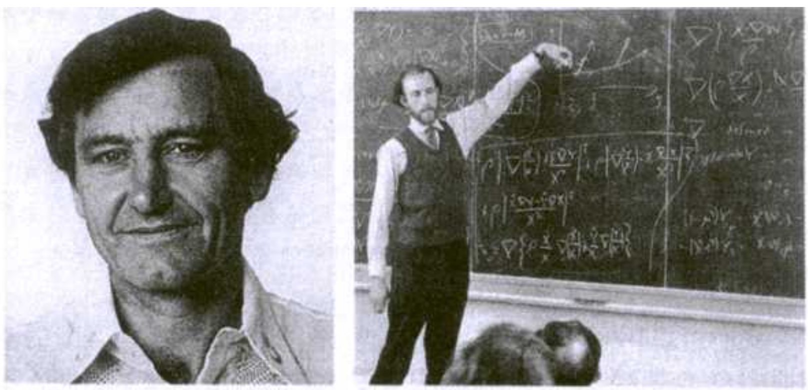
我们可以用嵌入图（图7.6左）来示意带电黑洞周围的空间曲率和黑洞的电力线。这幅图与图3.4的右下图基本相同，不过恒星没有了（图3.4的黑色部分），因为它在黑洞内部而不再和外部宇宙有联系。更严格地说，本图表现了黑洞外嵌在平直三维超曲面中的赤道“平面”——黑洞空间的一个二维碎片。（这个图的意义，请看图3.3和相关正文的讨论。）赤道“平面”是从黑洞视界切下来的，所以我们只能看到黑洞的外头，而看不到里面。视界实际上是一个球面，因为我们只看它的赤道，它在图中看起来就像一个圆。图表现了辐射状指向洞外的电力线。假如从上往下看（图7.6右），我们看不到空间曲率，但确实能把电力线看得更清楚。




图7.6　从带电黑洞视界发出的电力线。

左：嵌入图；右：俯视嵌入图

自旋对黑洞的影响，到60年代后期才为人们所认识，主要来自席艾玛在剑桥大学的一个学生，卡特尔。



左：R.克尔，约1975年。右：1972年6月B·卡特尔在法国阿尔卑斯山下的暑期讲习班上讲课。[左，R·克尔提供；右，K·索恩摄。]

卡特尔1964年秋参加席艾玛的小组时，席艾玛很快为他提出了第一个研究问题，建议他研究真实旋转恒星的坍缩。席艾玛解释说，以前关于坍缩的所有计算都是针对理想化的非旋转恒星的，现在对付旋转效应的工具和时机看来都成熟了。一个叫克尔（Roy Kerr）的新西兰数学家刚发表了一篇文章，给出了描述旋转恒星外的时空曲率的爱因斯坦场方程的解。[16]席艾玛解释说，这是从来没人发现过的第一个关于旋转恒星的解，但不幸的是，它是一个非常特殊的解，当然不能描述所有的旋转恒星。旋转的恒星有许多“毛”（许多性质，如复杂的

形状和气体的复杂运动），而克尔的解几乎没多少“毛”：时空曲率的形状很光滑，很简单，简单地对应于典型的旋转恒星。不管怎么说，克尔的爱因斯坦场方程解总还是一个起点。

很少有研究问题像这样立刻就有了收获：一年内，卡特尔就从数学上证明克尔解描述的不是旋转的恒星，而是旋转的黑洞。（伦敦的彭罗斯、利物浦的波耶（Robert Boyer）以及林凯斯特（Richard Linquist）——过去是惠勒的学生，现在在康涅狄格米德尔顿的卫斯理大学——都独立发现了这一点。）[17]到70年代中期，卡特尔和其他人又继续证明了，克尔解不仅描述了一类特殊的旋转黑洞，而且描述了所有可能存在的旋转黑洞。[18]

旋转黑洞的物理性质充实在克尔解的数学中，卡特尔通过认识数学而发现了那些性质应该是什么。[19]最有趣的是黑洞在它周围时空产生的龙卷风似的旋涡。




图7.7　黑洞旋转在空间产生的“龙卷风旋涡”的嵌入图。

旋涡如图7.7的嵌人图，喇叭型的曲面是嵌在三维平直超曲面内的黑洞的赤道面（黑洞空间的一个二维碎片）。黑洞的旋转抓住它周围的空间（喇叭曲面）像龙卷风一样跟着它旋转，速度与图中箭头的长度成正比。在远离龙卷风中心，空气旋转慢；同样，在远离黑洞视界，空间旋转慢。在龙卷风中心附近，空气旋转快；同样，在黑洞视界附近，空间旋转快。在视界处，空间紧紧附在视界面上，以完全相

同的速度跟它旋转。

空间的旋涡对落进黑洞的粒子的运动必然会产生影响。图7.8表现了两个这样的粒子径迹，图是在洞外静止观察者的参照系中看到的——也就是说观察者不落向视界进入黑洞。

第一个粒子（图7.8（a））下落很慢。如果黑洞不旋转，粒子将像坍缩恒星的表面那样，先是越来越快地呈辐射状向内运动，然后，在洞外的静止观察者看来，它将减慢下落速度，正好在视界上冻结。（回想一下第6章的“冻星”。）黑洞的旋转以非常简单的方式改变了这种情况：旋转在空间产生旋涡，空间旋涡带着靠近视界的粒子跟它一起旋转。于是，粒子被冻结在旋转的视界上，而在外面的静止观察者看来，它在绕着视界永不停息地旋转着。（同样，当旋转的恒星坍缩成旋转的黑洞时，外面的静止观察者会看到恒星表面被“冻结”在旋转的视界上，永不停息地绕着它转。）

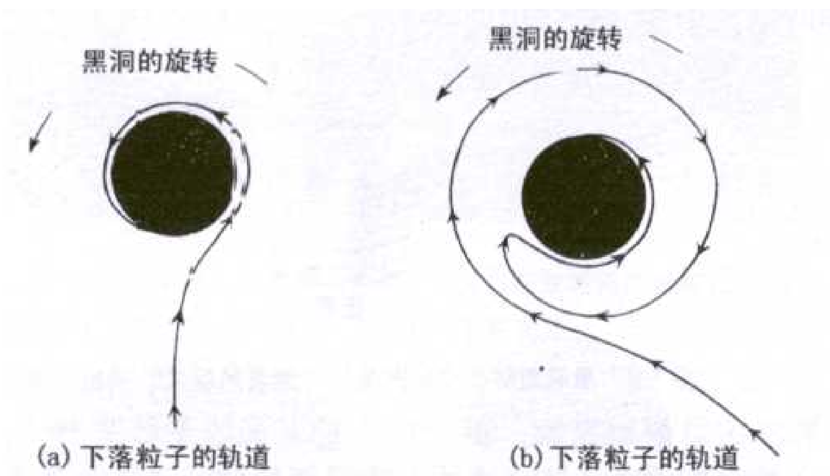



图7.8　落向黑洞的两个粒子在空间的轨道。（轨道是在远处的静止参照系中观测的。）尽管两个粒子的初始速度不同，但都被卷入空间旋涡，在靠近视界时随黑洞同步旋转。

虽然外面的观察者看到图7.8（a）的粒子冻结在旋转的视界上并永远留在那儿，但从粒子角度看情况就大不一样了。当粒子接近视界

时，由于引力的时间膨胀作用，粒子的时间与外面静止参照系的时间相比越来越慢。外面无限长的时间过去了，粒子却只经历了有限的很短的时间。在那个有限的时间里，粒子达到了黑洞的视界，在接下来的瞬间里，粒子将完全陷入视界，落进黑洞中心。下落粒子自己的经历与外面观察者看到的过程之间的这么巨大的差异，我们已经在恒星坍缩中遇到过了：从恒星表面看到的星体坍缩（快速陷入视界）与外面观察者看到的坍缩（坍缩冻结，见第6章最后部分）也是这样大不相同的。

第二个粒子（图7.8（b））是沿与黑洞旋转方向相反的螺旋轨道落下去的，但当它越来越靠近视界时，还是会被卷进空间旋涡，改变螺旋方向，在外面观测者看来，它也会像第一个粒子那样，被迫跟着视界旋转。

黑洞的旋转除了在空间产生旋涡，还会使视界产生变形，就像地球自转使地球表面产生变形一样。离心力会将旋转的地球赤道相对于两极向外拉出22千米。同样，黑洞的视界的赤道也会因离心力作用而像图7.9画的那样向外凸起。假如黑洞没有旋转，它的视界将是球形的（左图）。假如黑洞旋转很快，凸起就会很大（右图）。

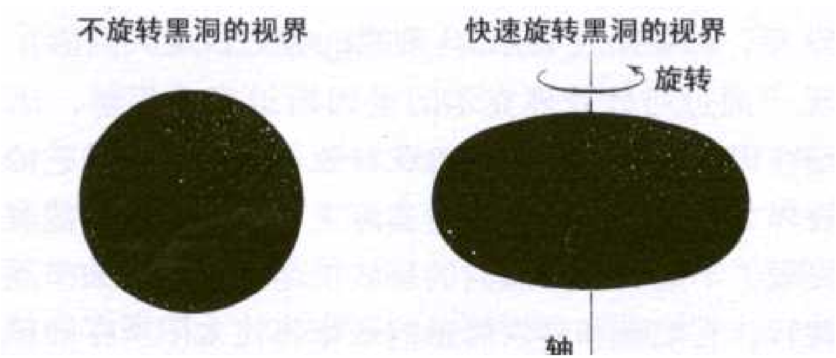



图7.9　两个黑洞视界的形状，一个不转（左），另一个以极大旋转速率的58%旋转。自旋对视界影响的结果是斯玛尔（Lany Smarr）1973年发现的，他是斯坦福大学的学生，受过惠勒的启发。

假如黑洞会急速地旋转，离心力会将视界撕裂开来，就像急速旋转的水桶将桶里的水抛洒出来。于是，存在某个能让黑洞存在的极大旋转速度。图7.9右边的黑洞在以这个极大值的58%的速度旋转。

那么，能让黑洞比允许的最大速度旋转更快一些吗？那样，它的视界会被破坏，而我们可以管窥它的内部。很遗憾，我们做不到。在1986年，也就是黄金年代过去10年后，伊斯雷尔证明了，不论用什么方法，想让黑洞比极大速度转得更快，总是要失败的。[20]例如，如果谁想通过向黑洞投进一些旋转更快的东西来提高以极大速度旋转的黑洞的速度，离心力会阻止快速旋转的物质到达视界进入黑洞。另外，也许急速旋转的黑洞与周围宇宙的小小随机相互作用（例如，远处恒星的引力作用）还会使旋转变慢一点。看来，物理学定律不想让黑洞外面的人偷看它的内部，不想让他发现藏在黑洞中心的奇点的量子引力秘密（见第13章）。

像太阳质量的黑洞，极大旋转速率是0.000　062秒（62微秒）一周。因为这个黑洞的周长大约是18.5千米，所以对应的转速是18.5千米/0.000 062秒，大约等于光速，299 792千米/秒。（这完全不是巧合！）质量为100万个太阳的黑洞，周长是太阳质量黑洞的100万倍，所以它的极大旋转速率（使它以光速转动的速率）就要小100万倍，即62秒转一圈。

1969年，彭罗斯（在第13章我们会更多地认识他）有一个惊人发现。[21]通过巧妙处理克尔的爱因斯坦场方程解，他发现旋转的黑洞在周围空间旋涡里藏着旋转能，因为旋涡和它的能量都在黑洞视界之外，所以这种能量实际上是可以用来作能源的。彭罗斯的发现了不起，因为黑洞的旋转能是巨大的。如果黑洞以极大速率旋转，它贮藏和释放能量的效率将比太阳所存的核燃料还高48倍。如果太阳在整个一生中燃尽全部核燃料（实际上是燃不尽的），也只有0.006的质量转化为光和热。如果去提取一个快速旋转黑洞的所有旋转能（这样会使它停下来），我们可以得到相当于48×0.006＝29%黑洞质量的可

利用能量。

奇怪的是，为了寻找一种大自然可能用来攫取和利用黑洞旋转能的可行办法，物理学家还费了7年的时间。他们找了一个又一个古怪的办法，在原则上都是可能的，但在最后发现自然的精灵前，没有哪个办法显得有多大的实用希望。我将在第9章讲这些寻找和发现，结果是，一个黑洞“机器”为类星体和巨大喷流提供了能源。

我们已经看到电荷产生指向黑洞视界外的辐射状电力线，旋转在黑洞周围空间产生旋涡，使黑洞形状发生扭曲，在旋涡中贮藏能量，那么，如果黑洞既有电荷又有自旋，会发生什么呢？不幸的是，答案没多大意思，什么新东西也没有。黑洞电荷产生通常的电力线，自旋产生通常的空间旋涡，贮藏旋转能量，使视界赤道以通常方式隆起。惟一新奇的事情是，空间旋涡在流过电场时会产生一些没多少意义的磁力线。（这些力线不是黑洞的新“毛”，不过是旧的标准形式的毛的相互作用的一个表现：自旋产生的旋涡与电荷产生的电场之间的相互作用。）旋转的带电黑洞的一切性质都体现在1965年得到的爱因斯坦场方程的一个精妙的解中，得到这个解的是皮兹堡大学的纽曼（Ted Newman）和他的一群学生：科赫（Eugene Couch）、金纳帕尔（K.Chinnapared）、埃克顿（Albert Exton）和托伦斯（Robert Torrence）。[22]

黑洞不仅能旋转，还会脉动。不过，脉动是旋转发现近10年后才在数学上发现的，发现途中曾遇到过一块巨大的拦路石。

惠勒的学生们“观察”了3年（1969~1971）的黑洞脉动，却不知道在看什么。这些学生有普赖斯（我的学生，也就是惠勒的徒孙）、维什维西娃拉（C.Y.Vishveshwara）和埃德尔斯坦（Lester Edelstein，马里兰大学米斯纳的学生，也是惠勒的徒孙），以及泽雷里（Frank Zerilli，惠勒自己在普林斯顿的学生）。他们借助计算机模拟和纸上计算观察黑洞的脉动。他们认为自己看到的是引力波（时

空曲率的波动），在黑洞附近反弹，又被黑洞自身的时空曲率捕获，但捕获不彻底，波动慢慢地从黑洞附近漏出来，然后飞散。这个想法不错，但没什么意思。

1971年秋，我的小组里的新研究生普雷斯意识到，时空曲率波动在黑洞附近的回弹可以想象为黑洞自身的脉动。毕竟，从视界以外看，黑洞除了时空曲率外什么也没有。所以，曲率的波动也只能是黑洞曲率的脉动，从而也就是黑洞本身的脉动。[23]

这一观点的改变产生了巨大影响。如果认为黑洞能够脉动，我们自然要问，它们与铃铛的脉动（“响铃”）或恒星的脉动有哪些相似？在普雷斯以前，没人问过这类问题；以后，这些问题就很显然了。

铃和星都有一个脉动的自然频率。（铃的自然频率产生纯粹的铃响音调。）那么，黑洞也有类似的自然脉动频率吗？是的，普雷斯通过计算机模拟发现了。这个发现激发了钱德拉塞卡和德维勒（Steven Detweiler，惠勒的学生的学生）开始对所有黑洞的自然脉动频率进行编目，在第10章，我们会回来谈这些频率，也就是黑洞的“铃声”。

快速转动的车轮稍微偏离正轨时，会产生振动，振动从旋转获得能量，越来越强烈。实际上，振动可以达到非常强烈的地步，在极端情况下，甚至可以使车轮脱离汽车。物理学家说“车轮振动是不稳定的”。普雷斯知道这一点，也知道旋转恒星有类似的现象。所以当他发现黑洞会脉动时，自然要问，“假如黑洞旋转很快，它的脉动会稳定吗？它们会从黑洞的旋转中获取能量，然后用这些能量来加强脉动吗？脉动会强烈到撕裂黑洞的地步吗？”钱德拉塞卡（那时对黑洞研究还不够深人）认为会的，我想不会。1971年11月，我们打了个赌。

判决输赢的工具那时还不存在。需要什么工具呢？由于脉动刚开始很弱，慢慢才会加强（如果能加强的话），可以认为它是黑洞时空

曲率的“微扰”——就像敲响的酒杯的振动，是对酒杯形状的小小扰动。这意味着黑洞的脉动可以用卡片7.1中大概介绍过的微扰方法来进行分析。不过，普赖斯、普雷斯、维什维西娃拉、钱德拉塞卡和其他一些人在1971年秋所用的具体的微扰方法却只适用于非旋转或旋转很慢的黑洞。为了快速旋转的黑洞的扰动，他们需要一种全新的微扰方法。

设计这种方法，成了1971年和1972年的热点问题。我的学生，米斯纳的学生，惠勒的学生以及钱德拉塞卡和他的学生弗里德曼（John　Friedman）像其他人一样都在为它工作。竞争很激烈，特奥科尔斯基赢了，[24]他是我的一个南非学生。



1972年12月在纽约Mama Kovdcs家的一次聚会。左起：K·索恩，M·普赖斯，B·普雷斯，R·特奥科尔斯基和S·特奥科尔斯基。[Sándor J.Kovács提供。]

特奥科尔斯基生动回忆了他那个微扰法的方程清晰出现的情这：“有时当你在搞数学的时候，你的头脑就开始模式选择了。”他说，“1972年5月的一个晚上，我正坐在我们在帕萨迪纳公寓的厨房的桌旁弄数学；妻子罗茜在特弗龙平底锅上煎薄饼，原想它不粘锅，但

还是粘了。每次她会倒些奶油面糊进去，在桌上敲锅，一边骂，一边敲，我大声喊她安静些，因为我正兴奋着；在我的数学公式中，一项项都相互抵消了，所有的东西都消失了！方程清晰出现了！我坐在那儿盯着那些惊人简单的方程，一心在想，我有多笨！6个月前就该发现它们了；我现在做的不过是把正确的东西加到一起。”[25]

用特奥科尔斯基方程，可以分析所有类型的问题：黑洞脉动的自然频率，黑洞脉动的稳定性，中子星被黑洞吞没时发出的引力波以及更多的其他问题。斯塔洛宾斯基（Alexi Starobinsky，泽尔多维奇的学生）、瓦尔德（Bob Wald，惠勒的学生）和科恩（Jeff Cohen，惠勒的学生布雷尔的学生）的研究小组和其他许多人很快就进行了这些分析，推广了特奥科尔斯基的方法。特奥科尔斯基和普雷斯在把握最重要的问题：黑洞脉动的稳定性。[26]

他们通过计算机计算和公式计算导出的结论令人失望：不论黑洞旋转多快，脉动都是稳定的。[1]黑洞脉动的确要从洞中攫取旋转能量，但它们也通过引力波辐射能量；而辐射能量的速率总是大于从黑洞的旋转中获得能量的速率。因此脉动能总会消失；脉动永远不会增强，从而黑洞不可能被脉动破坏。

因为普雷斯和特奥科尔斯基的结论关键靠的是计算机计算，钱德拉塞卡不满意，不承认打赌输了。只有直接通过公式做出了完整的证明，他才会完全信服。15年后，霍金以前的博士后怀庭（Bernard Whiting，也就是席艾玛的徒孙）给出了这样的证明，钱德拉塞卡认输了。[2]

钱德拉塞卡比我更喜欢完美。在完美主义者的名单上，他和泽尔多维奇各占一端。1975年，当黄金年代的年轻人宣布黄金年代结束了，所有的人都离开黑洞研究时，钱德拉塞卡很苦恼。这些年轻人掌握的特奥科尔斯基的微扰法足以证明黑洞可能是稳定的，但他们没有为这些方法带来恰当的形式，让其他物理学家能够自行计算任何可能

的黑洞扰动的所有细节——如脉动，来自下落中子星的引力波，黑洞炸弹或其他。他们的半途而废令人心痛。

于是，1975年，65岁的钱德拉塞卡把卓越的数学才能都用到了特奥科尔斯基的方程上。凭着用不完的精力和深刻的洞察，他穿过缤纷的数学丛林，将采撷的花朵织成一幅“洛可可式”的图画：[3]“华丽，壮观，赏心悦目。”1983年，他在73岁时终于完成了使命，发表了题为《黑洞的数学理论》的著作——这是未来10年黑洞研究者的数学手册，他们能想到的任何黑洞微扰问题都能从这本手册中找到解决的办法。[27]

[1]　芝加哥的德维勒和伊普色（James Ipser）独立提出了稳定性的一个有意义的数学证明；一年后，在圣巴巴拉加利福尼亚大学的哈特尔和维尔金斯（Dan Wilkins）又提出一个错误的证明。

[2]　钱德拉塞卡要为我订《花花公子》杂志，我那坚持男女平等的母亲和妹妹却令我感到惭愧，于是我要他为我订了《听众》。

[3]　rococo，原来的涵义是“贝壳形”，源自法语rocaille，也叫“路易十五式”，是法国路易十五时代（1715～1774）所崇尚的艺术，特征是具有纤细、轻巧、华丽和繁琐的装饰性。——译者

# 第8章　寻找

在天空寻找黑洞的方法提出来了，

追寻了，

也（大概）成功了

## 方法

想想你就是奥本海默。那是1939年，你刚让自己相信了大质量恒星在死后会形成黑洞（第5，6章）。那么，你会坐下来跟天文学家讨论在天空寻找黑洞存在证据的计划吗？不，完全不会。如果你是奥本海默，你的兴趣该在基础物理学上；你可能会把你的想法告诉天文学家，但你自己的注意力还在原子核——因第二次世界大战的爆发，它很快会将你卷入原子弹的设计。天文学家呢？他们会相信你的思想吗？不，一点儿也不。在天文学界，除了“野人”茨维基在推行他的中子星（第5章）外，普遍都很保守，世界到处都在反对钱德拉塞卡的白矮星极大质量（第4章）。

想想你就是惠勒，那是1962年，你刚克服了巨大的阻力，开始相信有些大质量恒星一定会在死后产生黑洞（第6，7章）。那么，你会坐下来跟天文学家讨论寻找它们的计划吗？不，完全不会。如果你是惠勒，令你感兴趣的是可能会在黑洞中心实现的广义相对论和量子力学那激动人心的结合（第13章）。你向物理学家们宣扬，星体坍缩的终点是出现新认识的转折点。你没有要天文学家们去寻找黑洞甚至中子星，关于寻找黑洞，你什么也没说；至于更有希望的找到中子星的

思想，你不过在文章里回应了天文学界的保守观点："这样的天体将有30公里量级的直径……会快速冷却下来……看这么暗淡的天体，跟看其他恒星的行星一样，几乎是没有希望的。"[1]（换句话说，一点希望也没有。）

想想你就是泽尔多维奇。那是1964年，你过去氢弹设计小组的成员波杜列茨刚在计算机上模拟了包括压力、激波、热、辐射和物质喷射等效应在内的星体坍缩（第6章），模拟产生了黑洞（或者说，计算机式的黑洞）。现在，你完全相信有些大质量恒星在死后一定会形成黑洞。那么，你会跟着坐下来和天文学家计划寻找它们吗？会的，当然会的。如果你是泽尔多维奇，你不大会同情惠勒为星体坍缩终点的困惑。[2]这个终点可能藏在黑洞的视界内，人们看不到它。但视界本身和黑洞对周围的影响却很有可能看到，你只需要指明如何去看。如果你是泽尔多维奇，你困惑的就是理解宇宙的可观测部分；那么，你怎么可能拒绝寻找黑洞的挑战呢？

寻找黑洞，该从哪儿开始？显然，始于我们的银河系——我们的$10^{12}$颗恒星的碟形集合。离我们最近的另一个大星系仙女嶝，有200万光年远，比银河系大20倍，见图8.1。这样，与银河系里相同的天体比，仙女座的恒星、气体云或其他天体将显得小20倍，暗400倍。于是，假如在银河系里都难探测黑洞，那么在仙女座，还要难400倍——在比仙女座还大10亿倍的星系里，那就更是难上难了。

既然在近处寻找那么重要，为什么不在我们的太阳系里找呢？也就是在从太阳到冥王星的范围内找？行星里不能有黑洞叫？不过因为它黑而我们看不见罢了。不，显然不会。如果真有黑洞，它的引力作用会比太阳还大；它会彻底破坏行星轨道，而我们没有发现这样的破坏。所以，最近的黑洞一定远在冥王星轨道以外。

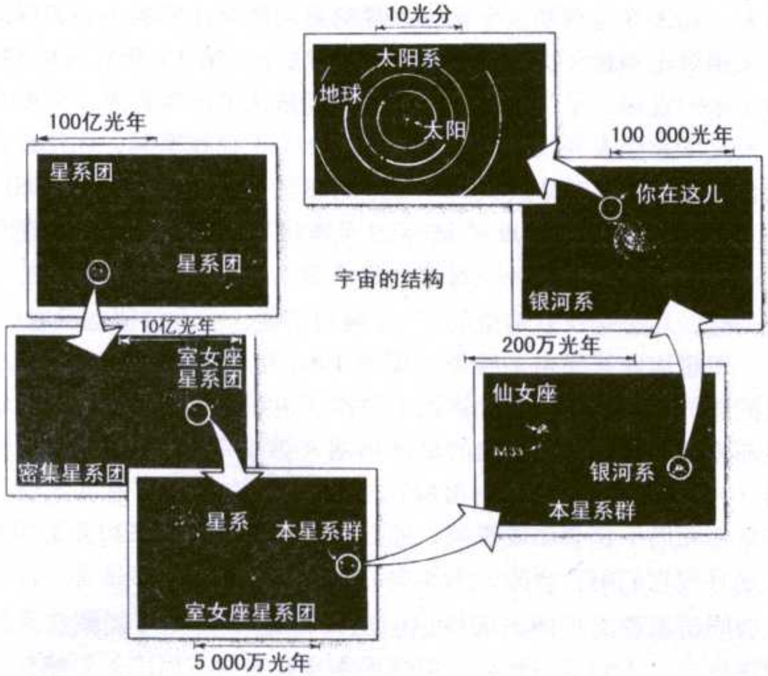



图8.1　宇宙结构简图

比冥王星远多少呢？你可以粗略估计一下。如果黑洞是大质量恒星死后形成的，那么最近的黑洞不太可能比最近的大质量恒星（天狼星）更近，离地球8光年；而且，它几乎肯定不会比（离太阳）最近的恒星半人马座α更近，4光年。

在这么远的距离上，天文学家怎么可能探测黑洞呢？他是不是在天空寻找一个运动的暗天体就行了？这样的天体可能会遮住来自它背后恒星的光。不行。因为黑洞的周长大概是50公里，而距离至少是4光年，它的圆盘张成的角不会大于$10^{-7}$弧秒，这大概相当于从月亮看一根人的头发的直径，比世界上最好的望远镜所能分辨的东西还小1 000万倍。运动的暗天体太小了，是不可能看到的。

如果说黑洞运动到恒星前面，我们看不见它的黑盘，那么能不能看到黑洞引力像透镜那样将恒星放大呢（图8.2）？恒星可能先显得

较暗，当黑洞运动到恒星和地球之间时它会亮起来，然后，黑洞走开了，它又暗下去，是这样吗？不，这样的寻找方法还是会失败的。原因是，恒星和黑洞可能会围绕彼此旋转从而靠得很近，也可能分开一个典型的星际距离。如果靠得近，那么小黑洞就像放在帝国大厦89层楼的窗台上的放大镜，然后，从几公里外来看它。当然，这个小小放大镜是没有能力放大一座大厦的；同样，黑洞对恒星的形状也不会产生什么影响。

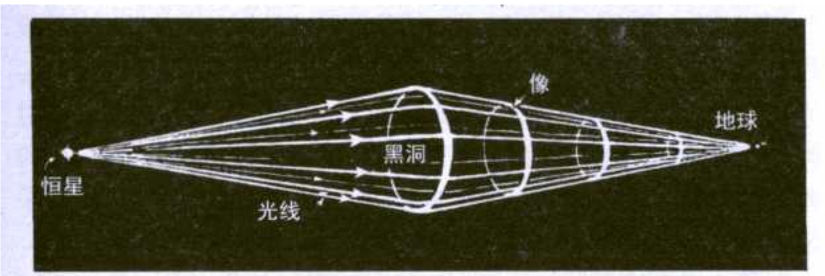



图8.2　从地球看，黑洞的引力作用像一个透镜，能改变恒星的外观大小和形状。图中的黑洞正好在恒星到地球的直线上，所以从恒星发出的光线将等量地从黑洞的上、下、前、后经过而到达地球。所有到地球的光线从恒星出发沿一个发散的锥体向外运动；经过黑洞后，光线发生弯曲，然后沿一个收敛的锥体到达地球。结果，在地球的天空上，恒星的像是一个细环，与没有黑洞时的像相比，它的表面积大得多，从而也亮得多。望远镜分辨不出这么小的环，但恒星的总亮度可以提高10-100倍，或更多。

如果恒星和黑洞像图8.2那样分得很远，但聚焦能力还是很强，可以将恒星亮度提高10～100倍。然而，星际距离总是巨大的，要地球-黑洞-恒星处在一条直线上，是极其难得的事情，不会有希望。而且，即使观察到了这样的透镜，从恒星到地球的光经过黑洞要经历很长的距离（图8.2），在这么长的距离里可能会有一颗恒星处在黑洞的位置，起着透镜的作用。所以，地球上的天文学家不可能知道，透镜是一个黑洞，还是仅仅是一颗普通而暗淡的恒星。

泽尔多维奇在寻找观测黑洞的方法时，一定进行过像这样的一连串的推理。他最后得到一个有希望的方法（图8.3）：假设黑洞和恒星在围绕彼此的轨道上（形成双星系），天文学家在将望远镜对准这

对双星时，只能看到来自恒星的光，黑洞是看不见的。不过，恒星的光将提供黑洞存在的证据：恒星在轨道上绕黑洞运动时，先朝着地球，然后离开地球。当它朝向我们时，多普勒效应将使星光向蓝色移动；当它远离我们时，光会红移。天文学家可以高精度地测量这些频移，因为星光经过摄谱仪（一种精巧的棱镜）时会显出尖锐的谱线，这种谱线的波长（颜色）的轻微移动都会清楚地表现出来。通过波长移动的测量，天文学家能推测恒星走近或离开地球的速度，通过测量频移随时间的变化，他们能推测恒星的速度如何随时间变化。速度变化的典型大小约在每秒10～100千米之间，测量的典型精度是0.1千米/秒。

从这样高精度的恒星速度测量，能了解些什么呢？我们会了解一些关于黑洞质量的东西：黑洞质量越大，作用在恒星上的引力就越强，从而阻止恒星被吸进黑洞所要求的离心力也越强。为获得强大的离心力，恒星必然在轨道上快速运动。因此，大的速度总是伴随着大的黑洞质量。

那么，为了寻找黑洞，天文学家应该找那些光谱表现出周期性的红—蓝—红—蓝频移的恒星。这类移动是恒星有一颗伴星的确凿证据。天文学家应该测量恒星的光谱，从它推测恒星绕伴星运动的速度，然后从速度推测伴星的质量。如果伴星质量很大，而且没有发出一点光，那么它很可能是一个黑洞。这就是泽尔多维奇的建议。

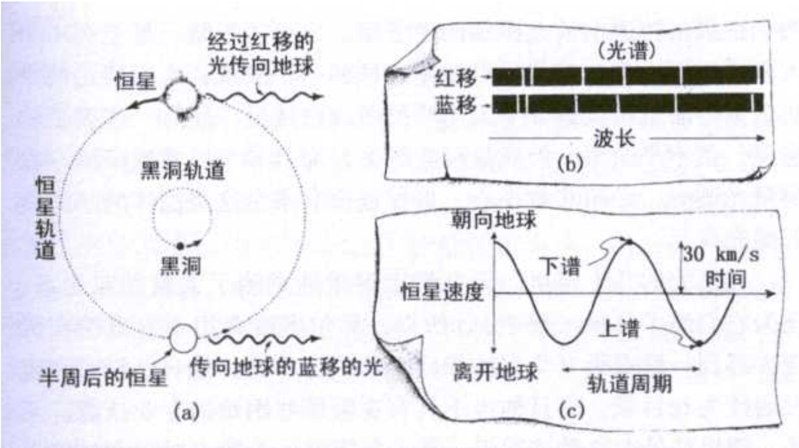



图8.3　泽尔多维奇提出的寻找黑洞的方法。（a）黑洞与恒星在相互环绕的轨道上。如果黑洞比恒星重，它的轨道就会像图中那样比恒星的小（就是说，黑洞动一点，而恒星动很多）。如果黑洞比恒星轻，它的轨道会更大（就是说，恒星动一点，而黑洞动很多）。当恒星背离地球运动时，光会向红端移动（向更长的波长）。（b）进入地球望远镜的光通过摄谱仪形成光谱。图中画了两个谱，上面记录的是恒星背离地球时的谱线，下面记录的是半个周期后恒星朝向地球运动时的谱线。谱中细线的波长有相对移动。（c）通过测量这样的光谱序列，天文学家能确定恒星朝向和背离地球的速度如何随时间变化；根据这个变化的速度，他们可以确定恒星所围绕的天体的质量。如果质量大于2个太阳且没有光从它发出，那这个天体可能就是一个黑洞。

尽管这个方法比以前的任何一个都好得多，但它还是隐藏着许多问题，我只谈两个：第一，暗伴星的质量不是直接“称”出来的。恒星的测量速度不仅依赖于伴星的质量，而且依赖于它自身的质量和双星系轨道面对我们视线的倾角。恒星质量和轨道倾角倒可以通过仔细的观测推想出来，但这么做不容易，精度也不会高。结果，估计的暗伴星的质量很容易出现大误差（比如，差2倍或3倍）。第二，黑洞不是恒星能有的惟一暗伙伴。例如，中子星伴星也可能是暗的。为了肯定伴星不是中子星，需要确认它比中子星所能允许的最大质量（约2个太阳质量）大得多。两个在紧密轨道上彼此围绕的中子星，也可能很暗，甚至有4个太阳质量那么重。暗伴星可能是这样的一个系统；也可能是两个轨道紧密的总质量达3个太阳质量的冷白矮星。还有一些类型的恒星，虽不是全黑，但质量可能很大，而且异乎寻常地暗淡。在测量光谱时，必须非常小心，以保证没有来自这类星体的一丝微弱

的光亮。

在过去的几十年里，天文学家艰难地观测了大量的双星系，还为它们编了目录（星表），所以，泽尔多维奇用不着直接在天空去寻找，检索天文学家的编目就够了。不过，他自己没时间也没耐性去查目录，而且他也不具有克服那些困难的专业技能。于是，跟以往处在这种境况下一样，他需要一个能在时间和技能上帮助他的人——这回是古赛诺夫（Oktay Guseinov），一个天文学研究生，对双星已经很熟悉了。他和泽尔多维奇一起，在星表的几百个数据充分的双星系中找到了5个有希望的黑洞候选者。[3]

接下来的几年里，天文学家几乎没有谁注意那5个黑洞候选者。我对天文学家的漠不关心感到很苦恼，于是在1968年，我请加州理工学院的天文学家特里姆布尔（Virginia Trimble）来帮我修订并扩充泽尔多维奇-古赛诺夫黑洞候选者名单。虽然特里姆布尔刚读完博士，但在天文学领域的专业知识已经很丰富了。她知道我们可能会遇到的所有问题——除上面说的，还有更多——她能准确地估价它们。我们自己在双星目录中搜寻，用所有能找到的公开数据校核最有希望的双星，最后得到一张有8个黑洞候选者的名单。[4]不幸的是，对这8种情况，特里姆布尔都能提出一个半推理的非黑洞的解释，说明这个伴星为什么那么黑。今天，四分之一世纪过去了，我们的候选者没有一个留下来的，现在看来，它们似乎没有一个是真正的黑洞。

泽尔多维奇在考虑寻找黑洞的双星方法时，知道那是一场赌博，没有一点儿赢的把握。幸运的是，他灵机一动，又有了第二个想法——那是1964年，纽约绮色佳康奈尔大学的天体物理学家萨尔皮特（Edwin Salpeter）也同时独立地发现了这个想法。[5]

假设一个黑洞穿过气体云——或者等价地说，在黑洞看来，气体云通过它（图8.4），被引力加速到近光速的气流将在黑洞上下绕着它飞流，涌向黑洞的边缘。涌来的气体形成激波前沿（密度突然大

增），将气体的巨大能量转化为热，引起强烈辐射。于是，从结果看，黑洞的作用像一台机器，将一些下落气体的质量转化为热量，然后辐射。这台“机器”的效率可能很高，多高呢？泽尔多维奇和萨尔皮特得到的结果是，比核燃料燃烧的效率还高。

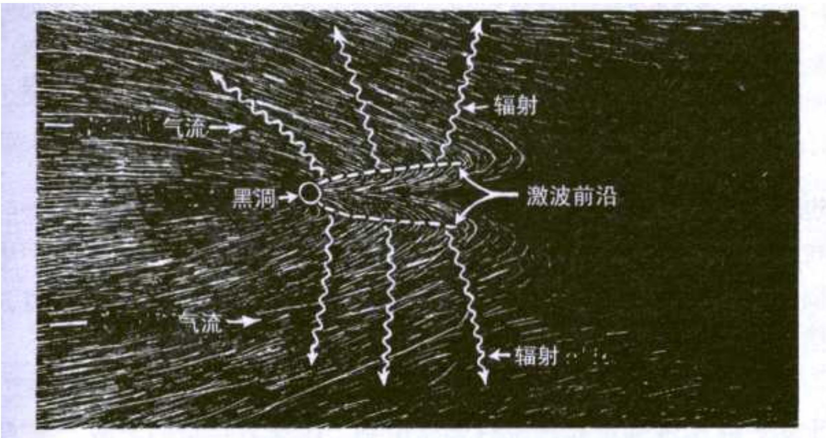



图8.4　萨尔皮特-泽尔多维奇寻找黑洞的建议。

对这个问题，泽尔多维奇和他的小组研究了两年，他们从不同角度来看它，想寻找一种办法让它更有希望。不过，这只是一个问题，没有引起多少注意；关于黑洞、中子星、超新星和宇宙起源，他们还有很多的想法。于是，1966年的某一天，在一场激烈争论中，泽尔多维奇和诺维科夫一起认识到，他们可以将双星思、想与下落气体的思想联合起来（图8.5）。[6]

强大的气体（大多是氢和氦）风吹散部分恒星表面。（太阳也吹这类风，不过很弱。）假定黑洞和刮风的恒星在相互环绕的轨道上，黑洞将捕获一些气体，在激波前沿将它加热，使它产生辐射。泽尔多维奇在他莫斯科家里1米见方的黑板上同诺维科夫估算了被激气体的温度：几百万度。

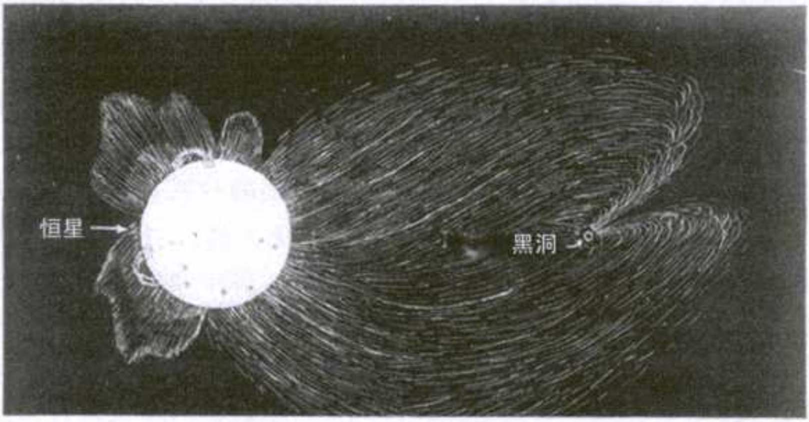



图8.5　泽尔多维奇和诺维科夫提出的黑洞找寻办法。黑洞引力捕获了那些将伴星表面吹散的风。风的气流从相对的方向包围黑洞，碰撞形成尖锐的激波前沿，气体在这儿被加热到几百万度，并发出X射线。光学望远镜可以看到绕着一个大质量的暗伴星的恒星，X射线望远镜能看到来自伴星的X射线。

这样高温的气体不会发出太多的可见光，而是发出紫外线。于是，泽尔多维奇和诺维科夫认识到，围绕着伴星的黑洞，有些（虽然不是大多数）可能会发出大量X射线。

因此，为了找黑洞，可以将光学望远镜与X射线结合起来。黑洞候选者所在的双星，一个是光学的亮星，X射线的暗星；另一个应该是在光学上暗淡而在X射线上明亮的天体（黑洞）。由于中子星也能从伴星捕获气体，在激波前沿将它加热，产生X射线，所以确定光学暗淡而X射线明亮的天体的质量是决定性的，我们必须确信它比2个太阳重从而不会是中子星。

这样的寻找策略还有一个问题。1966年，X射线望远镜还太原始了。

## 寻

假如你是天文学家，X射线令你烦恼的地方在于，它们不能穿透地球大气。（对人来说，这是好事，因为X射线会诱发癌症和变异。）

幸运的是，富于幻想的实验物理学家们在美国海军研究实验室（NRL）的弗里德曼（Herbert Friedman）领导下，从40年代就开始在为空间X射线天文学打基础了。二战后不久，弗里德曼就和他的伙伴们用缴获的德国V-2火箭发送仪器研究太阳。弗里德曼描述过1946年6月28日的第一次飞行。他们在火箭头上装一台摄谱仪，探测太阳的远紫外辐射。（跟X射线一样，远紫外辐射也不能穿透地球大气。）经过短时间大气上空飞行，收集数据后，“火箭返回地球，头朝下，流线式地飞落下来，埋进一个直径约80英尺[24米]、深30英尺[9米]的大坑，挖了几个星期才挖出一堆难以识别的碎片，火箭仿佛在坠落时‘蒸发’了。”[7]

从这个不幸的结局开始，弗里德曼等人凭他们的创造力、坚韧精神和努力工作，终于将紫外线和X射线天文学一步步建立起来了。1949年，弗里德曼和他的同事们用V-2火箭载盖革计数器上天探测来自太阳的X射线。50年代后期，他们的计数器搭上了美国造的高空探测火箭，不仅探测太阳辐射，也探测恒星辐射。不过，X射线是另一回事。太阳的X射线每秒钟落到计数器的数量是每平方厘米100万，所以探测太阳的X射线相对要容易一些。但是，据理论估计，最亮的X射线星也比太阳暗10亿倍，探测这么暗淡的一颗星，需要的探测仪应比弗里德曼1958年用的灵敏1 000万倍，这个要求很高，但不是不可能。

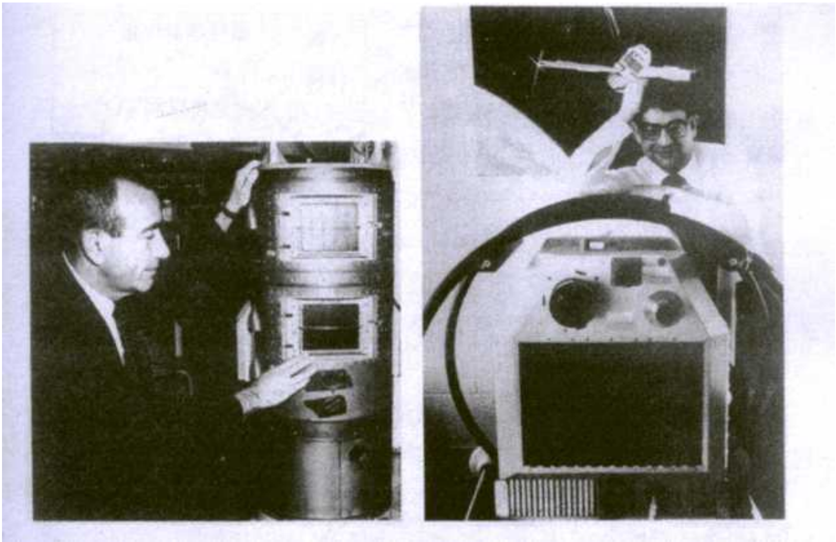
到1962年，探测器提高了10 000倍，就等后来的1 000倍了。这

时，在弗里德曼的进展激发下，其他研究小组也来竞争了。其中，贾柯尼（Riccardo Giacconi）领导的一组将成为最有力的对手。

贾柯尼的成功，泽尔多维奇也可以说有特别的一份功劳。1961年，苏联令人意外地废除了苏美两国3年不搞核武器试验的协议，进行了一次人类最大威力的原子弹试验——就是泽尔多维奇和萨哈洛夫的小组在基地设计的那种原子弹（第6章）。美国恐慌了，也准备搞自己的新试验。这将是地球轨道飞船时代美国的第一次试验，也是第一次有可能在空间探测核爆炸放出的X射线、y射线和高能粒子，这些观测对监测苏联未来的核试验是很重要的。不过，为了观测即将进行的一系列试验，美国还需要一个紧急计划。组织和领导这个计划落到了28岁的贾柯尼头上，他是美国科学与工程公司（一家在剑桥、麻省的私营公司）的实验物理学家，刚开始设计弗里德曼那样的X射线飞行探测器。美国空军为贾柯尼提供了他需要的经费，但时间很紧。在不到1年的时间里，他的6人X射线天文学小组壮大到了70人，他们设计、制造并试验了多种核武器爆炸探测仪，进行了24次火箭和6次卫星飞行，成功率达95%。这些经历将他的小组铸成了一支忠诚的、有奉献精神和熟练技能的队伍，他们武装好了自己，能同建立X射线天文学的所有竞争者战斗。

贾柯尼老练的小组在天文学迈出的第一步，是用弗里德曼式的探测仪，像弗里德曼那样让它搭上探测火箭，寻找来自月球的X射线。1962年6月18日午夜前1分钟，火箭从新墨西哥州的白沙基地升空，然后落回地球。火箭在地球大气层外飞行了350秒，高度足以探测月球的X射线。发回地球的数据令人疑惑：X射线比预想的强得多。仔细检查这些数据，就更令人惊讶了。X射线似乎并不来自月球，而是来自天蝎座（图8.6（b））。贾柯尼和他的伙伴（古尔斯基（Herbert Gursky）、鲍里尼（Frank Pa-olini）和罗西（Bruno Rossi））用了两个月的时间来寻找数据和仪器中的错误，但什么也没找到。然后，他们宣布了自己的发现：人类所探测的第一颗X射线星比理论天体物

理学家所预言的亮5 000倍。[8]10个月以后，弗里德曼的小组证实了这一发现。这颗星被命名为ScoX-1（天蝎座X-1，1代表“最亮”，X指“X射线源”，Sco即“天蝎座（Scorpius）”的缩写）。



左：H·弗里德曼和火箭上的探测仪，1968年。右：R·贾柯尼和自由号X射线探测仪，约1970年。[左，美国海军实验室提供；右，贾柯尼提供。]

理论家哪儿错了呢？他们怎么会把宇宙的X射线强度低估了5 000倍呢？原来，他们错误地假定，X射线天空的主要天体也就是我们已经熟悉了的光学天空中的天体——如月球、行星和正常恒星，而它们实际上却是很弱的X射线源。不过，天蝎X-1和即将发现的其他X射线星并不是以前人们见过的那些天体，它们是中子星和黑洞，从正常伴星那儿捕获气体，然后将它们加热到很高的温度；至于加热方式，泽尔多维奇和诺维科夫很快会提出来（上面的图8.5）。然而，为了确认这的确就是所观测到的X射线星的本质，弗里德曼和贾柯尼那样的实验家与泽尔多维奇和诺维科夫那样的理论家还得并肩奋斗10年。

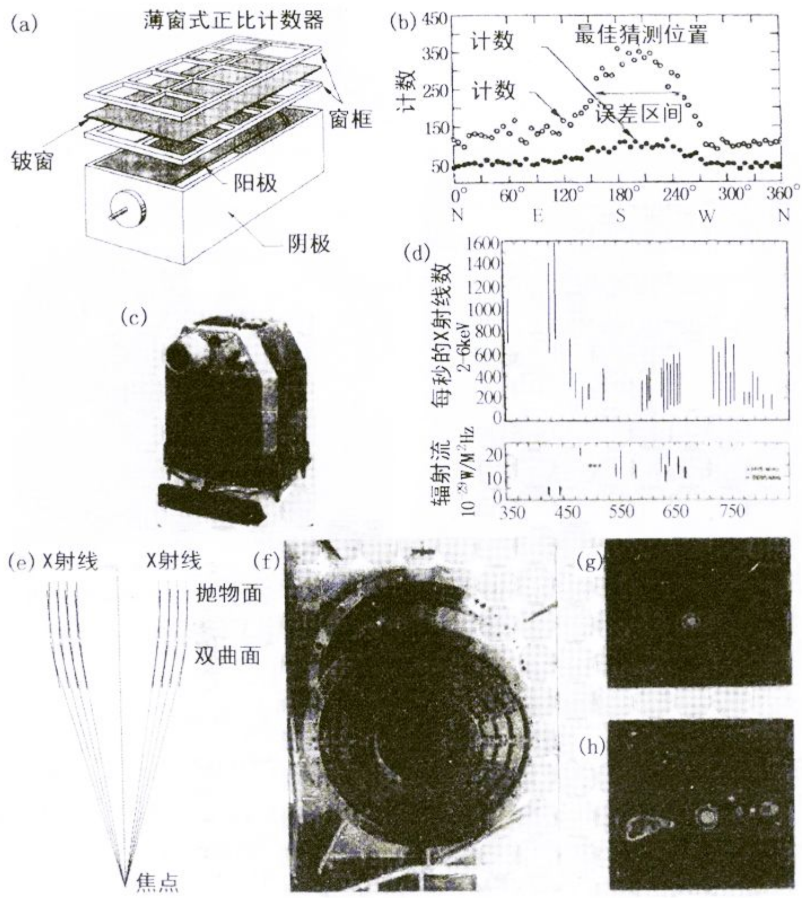



图8.6　从1962年到1978年，X射线天文学仪器的技术和结果的进步。（a）贾柯尼小组1962年用以发现第一颗X射线星的盖革计数器略图。（b）盖革计数器的数据，说明那颗星不在月球的位置。注意角分辨率太低（误差区间太大），达90°。（c）1970年自由号X射线探测仪：盒内装有经过巨大改进的盖革计数器，计数器前有扇百叶窗，只有近垂直到达窗口的X射线才可能测到。（d）自由号测得的来自候选黑洞天鹅X-1的X射线。（e）和（f）是1978年爱因斯坦X射线望远镜的X射线聚焦镜头的示意图和照片。（g）、（h）是爱因斯坦望远镜拍摄的两个候选黑洞的照片：天鹅X-1和SS-433。[贾柯尼私人照片和图片。]

贾柯尼1962年用的探测器极简单（图8.6（a）），就是一个顶上有扇薄窗户的充电气室。X射线穿过窗户进入气室后，会从气体的原子中击出一些电子；这些电子被电场吸到一条线上，在那儿产生电流，这就意味着X射线来了。（这种气室有时被称为盖革计数器，有

时又被称为正比计数器。）载着气室的火箭每秒钟自转两圈，箭头慢慢从上转到下，气室窗也跟着扫过大片天空，从一方指向另一方。当它指向天蝎座时，会记录到很多X射线；指向别处时，记录会很少。不过，X射线可以从很多方向进入窗口，所以气室估计的天蝎X-1在天空的位置很不确定，它只能报告一个猜测的最佳位置。从它那90°的误差区间可以想象，这个最佳的位置错到哪儿去了（图8.6（b））。

为了确定天蝎X-1和其他很快发现的X射线星实际上就是双星系中的中子星和黑洞，误差区间（在天空中位置的不确定性）必须在几弧分以内，这个要求很高，将角精度提高了1 000倍。

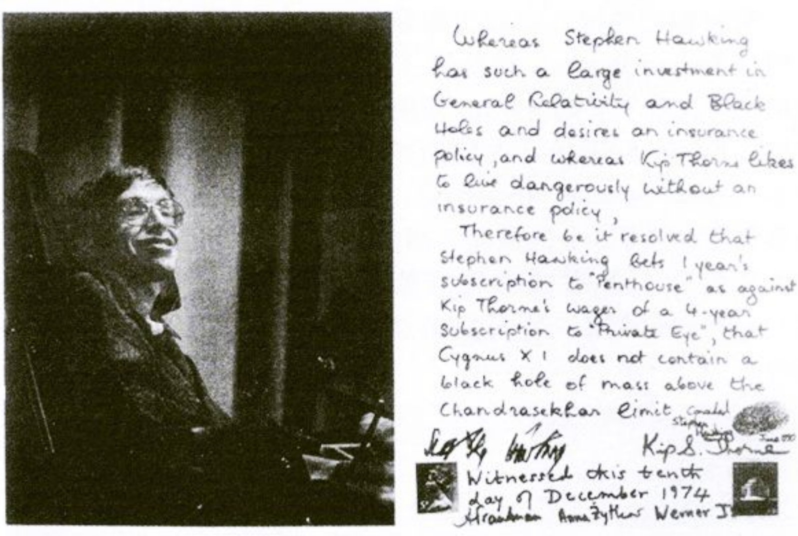
人们盼望的更大的进步，在接下来的16年里，在几个小组（弗里德曼的、贾柯尼的和其他的）竞争中，一步步实现了。当一系列不断改进的探测器一个跟着一个装上火箭在天空飞过之后，1970年12月，第一颗X射线探测卫星自由号也上天了（图8.6（c））。贾柯尼小组制造的这颗自由号，装着一个充满气体的X射线计数器，比他们1962年装在火箭上的那个大100倍。气室的窗户有百叶帘式的板条，可以阻挡一些X射线，从而气室只能探测到从垂直方向和附近几度来的X射线（图8.6（d））。自由号发现并编目了339颗X射线星，后来，美国、英国和荷兰科学家又制造了几颗类似的但更特殊的X射线卫星。到1978年，贾柯尼小组继自由号后又放飞了一颗大卫星，爱因斯坦号，这才是世界上第一个真正的X射线望远镜。因为X射线会穿透它直射在上面的物体，包括镜片，所以爱因斯坦望远镜用了一组安置好的镜片，X射线将像雪橇在冰雪坡面上滑行那样溜过这些镜片（图8.6（e），（f）），它们将X射线聚焦成天空中1弧秒大小的像——像的精度与世界上最好的光学望远镜一样（图8.6（g），（h））。

从贾柯尼火箭到爱因斯坦望远镜经过的短短16年（1962～1978）时间里，角测量精度提高了300 000倍，在这个过程中，我们对宇宙的认识也发生了革命性的变化：X射线透露了中子星和可能的黑洞；通过X射线，我们发现了对流的热气体，星系就沐浴在巨大星

系团的这种气体中；我们还看到了超新星残骸中的热气体，某些恒星冕（恒星外部的大气）里的热气体以及星系和类星体核中的超高能粒子。

在X射线探测器和X射线望远镜发现的几个黑洞候选者中，天鹅X-1（简写为Cyg X-1）是最可信的一个。1974年，它刚成为最可能的候选者后不久，霍金和我打了个赌，他赌它不是黑洞，我赌它是。

打赌10年后，我和卡洛丽（Carolee　Winstein）结婚了，她为赌注感到很难堪（如果我赢了，我将得到《阁楼》（Pent-house）杂志；如果史蒂芬赢了，将得到《私家侦探》（Private　Eye）杂志），我的姐妹和母亲也觉得讨厌。[1]但她们用不着担心我会真赢得1年的《阁楼》（在80年代我是这么想的）；关于天鹅X-1的本质，我们认识的进步还很慢。到了1990年，在我看来，也只有95%的把握相信它是一个黑洞，这还不足以令史蒂芬认输，显然，他对证据有不同认识。后来，1990年6月的一个晚上，我正在莫斯科与前苏联同行工作时，史蒂芬带着家属、护士和朋友闯进我在加州理工学院的办公室，把赌约找出来，在上面签字画押，认输了，还印上一只大拇指印。[2]


Whereas Stephen Hawking has such a large investment in General Relativity and Black Holes and desires an insurance policy, and whereas Kip Thorne likes to live dangerously without an insurance policy,

Therefore be it resolved that Stephen Hawking bets 1 year's subscription to "Penthouse" as against Kip Thorne's wager of a 4-year subscription to "Private Eye", that Cygnus X1 does not contain a black hole of mass above the Chandrasekhar limit.

Conceded Stephen Hawking June 1990

Kip S. Thorne

Witnessed this tenth day of December 1974

Anna Zytkow Werner Israel

右：史蒂芬·霍金和我为天鹅X-1是否是黑洞而立的赌约。左：1990年6月，霍金在南加利福尼亚大学演讲，两小时后他就闯到我的办公室在赌约上签字认输了。[霍金照片由南加利福尼亚大学Irene Fertik提供。]

天鹅X-1包含着黑洞的证据，也就是泽尔多维奇和诺维科夫提出黑洞找寻方法时想象的那种：天鹅X-1是由一颗光学明亮而X射线暗淡的恒星围绕一颗X射线明亮而光学暗淡的伴星组成的双星系统，那颗伴星的质量已经肯定比中子星大得多，从而可能是一个黑洞。

天鹅X-1本性的证据却不是那么容易发现的，全世界数以百计的实验物理学家、理论天体物理学家和实测天文学家从60年代到70年代团结协作，为它付出了大量的努力。

实验物理学家，如弗里德曼、鲍耶尔（Stuart Bowyer）、拜拉姆（Edward Byram）和查伯（Talbot Chubb），在1964年通过火箭飞行发现了天鹅X-1；塔南鲍姆（Harvey Tananbaum）、克罗格（Edwin Kellog）、古尔斯基、马雷（Stephan Marray）、施莱尔（Ethan Schrier）和贾柯尼，1971年用自由号将天鹅X-1的位置确定到2弧分的误差区间内（图8.7）；另外还有许多人，发现和研究了X射线和它的能量的剧烈而混乱的波动[涨落]——这些涨落，是人们希望黑洞周围的湍流热气体应该有的。

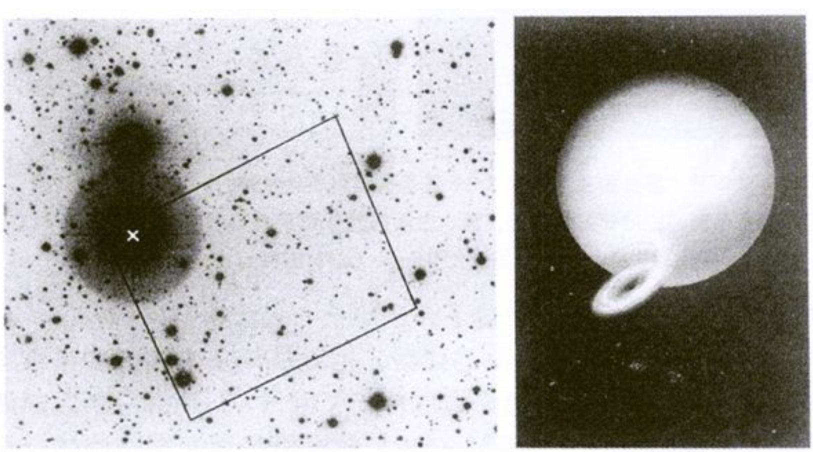


图8.7　Jerome Kristian 1971年用帕洛玛山的5米光学望远镜拍摄的一张照片负片。黑色矩形框是自由号在1971年确定的天鹅X-1位置的误差区间。白色×是射电望远镜观测到的无线电波爆发，与天鹅X-1的X射线的突然变化是一致的。×的位置正好是光学恒星HDE 226868，所以认定它是天鹅X-1的一颗伴星。1978年，爱因斯坦X射线望远镜证实了这一点，见图8.6g。右图是艺术家根据所有光学和X射线数据描绘的天鹅X-1和HDE 226868。[左：卡内基天文台Jerome Kristian博士摄；右：Victor J.Kelley绘，国家地理学会提供。]

投身进来的全世界的实测天文学家，如热尔曼（Robert Hjellming）、瓦德（Cam Wade）、布雷斯（Luc Braes）和米莱（George Miley），1971年在自由号天鹅X-1的误差区间内发现了一次无线电波爆发，同时还伴随着自由号测得的天鹅X-1的X射线的巨大变化，于是，天鹅X-1的位置被确定到1弧秒内（图8.6（d）和8.7）；韦伯斯特（Louise Webster）、默丁（Paul Murdin）和波尔顿（Charles Bolton）用光学望远镜发现，在无线电波爆发处的一颗光学恒星，HDE 226868，绕着一颗大质量的光线暗淡而X射线明亮的伴星（天鹅X-1）；另外，大约上百的光学天文学家对HDE 226868和它附近的恒星进行了艰难的测量，为了避免天鹅X-1的质量估计出现严重错误，这些测量是很重要的。

投身进来的理论天体物理学家包括泽尔多维奇和诺维科夫，他们提出了黑洞的找寻方法；帕津斯基（Bohdan Paczynski）、阿伏尼（Yoram Avni）和巴考尔（John Bahcall），他们发现了一种复杂但可靠的方法来避免质量估计的错误；布尔比基（Geoffrey Burbidge）和普伦德加斯特（Kevin Prendergast），他们认识到X射线发出的热气体会在黑洞周围形成盘；沙库拉（Nikolai Shakura）、桑尼耶夫（Rashid Sunyaev）、普林格尔（James Pringle）、里斯、奥斯特里克（Jerry Ostriker）以及其他许多人，为了与X射线观测对比，发展了X射线发出的气体和它形成的盘的理论模型。

通过这些巨大的努力，到1974年，我们对天鹅X-1和它的伴星HDE 226868的图景有了80%的信心，图8.7的右边是艺术家为它们画的像，其实那正是泽尔多维奇和诺维科夫想象的，不过要具体得多：天鹅X-1中心的黑洞质量肯定大于3个太阳质量，也许大于7个太阳，

最可能是16个太阳。它的光线明亮而X射线暗淡的伴星HDE 226868大约有20个太阳以上的质量，很可能是33个太阳，而它的半径比太阳约大20倍，它的表面到黑洞的距离约为20个太阳半径（1 400万公里），双星距地球约6 000光年。天鹅X-1是X射线天空中的第二亮星；HDE 226868尽管比大多数望远镜看到的恒星更亮，但对肉眼来说，还是太暗了。

自1974年以来的近20年中，我们对天鹅X-1图景的信心从大约80%提高到了95%（这是我个人的估计）。我们还不能100%地相信它，因为尽管费了那么大的努力，但还没有在天鹅X-1中发现黑洞的确凿信号。没有任何X射线的或光的信号向天文学家坦白，“我来自黑洞。”所有这些观测，都还可能找到别的非黑洞的解释，尽管那些解释都太拐弯抹角，没几个天文学家当真。

反过来，一些中子星（被称为脉冲星）倒发出了确凿的“我是中子星”的呼喊：它们的X射线，有时也可能是无线电波，产生了非常准时的尖锐脉冲。脉冲时间有时跟我们最精确的原子钟一样精确。这些脉冲只能解释为从中子星表面射出的辐射流，随星体旋转而扫过地球——像机场或灯塔的闪光灯光。为什么这是惟一可能的解释呢？如此精确的定时只能来自大质量天体的旋转，它有大的惯性，从而对产生不规则时间的不规则力量有大的抵抗能力。在天体物理学家头脑中想象的所有大质量天体中，只有中子星和黑洞能以某些脉冲星那样大的速率（每秒几百圈）旋转；而又只有中子星能产生旋转的辐射束，因为黑洞是没有“毛”的。（任何这样的固定在黑洞视界的辐射束也是一类“毛”的例子，但黑洞却留不住它们。[3]）

为了找到像脉冲星的脉冲那样的黑洞的确凿信号，天文学家在天鹅X-1中寻找了20年，但什么也没找到。这类信号的一个例子是（泽尔多维奇小组的桑尼耶夫在1972年提出的），[9]在围绕黑洞旋转的凝聚气团中出现的旋转束所产生的脉冲星式的辐射脉冲。如果气团离黑洞近，它能维持在许多轨道上，直到最终落入黑洞视界，这时候，它

的脉冲之间逐渐变化的时间间隔就可能清楚而确凿地表示“我是一个黑洞”。不幸的是，这种信号从来没人见过，原因似乎有几个：（1）X射线发出的热气体绕黑洞的运动混乱无常，凝聚的气团只能在一个或几个轨道上维持。（2）假如几个气团真能长时间聚集在一起并产生黑洞信号，其余的混沌气体产生的混乱X射线显然也会将那些信号淹没。（3）假如天鹅X-1真是一个黑洞，那么根据数学模拟，大多数X射线应该来自远离视界的地方——也就是周长大于10倍临界值或更远的地方，那儿比视界附近有更大的发射X射线的空间。在距黑洞这么远的地方，广义相对论和牛顿引力理论的预言是近似相同的，所以，即使有来自轨道气团的脉冲，也不会带来强大的确定的黑洞信号。

由于同样的一些原因，天文学家也许永远不可能从黑洞附近产生的电磁波中发现任何类型的确定的黑洞信号。幸运的是，我们有很大的希望看到一类完全不同的黑洞信号：引力辐射所携带的信号。等到第10章我们再来谈。

黑洞理论研究的黄金年代（第7章）也伴随着寻找黑洞的观测研究，也在那个年代，发现了天鹅X-1，阐明了它的本质。于是，可能有人以为曾拥有黄金年代的那些年轻人（彭罗斯、霍金、诺维科夫、卡特尔、伊斯雷尔、普雷斯、特奥科尔斯基、普赖斯，等等）在寻找黑洞中也会担起里要角色，实际并非如此，只有诺维科夫例外。他们显露的才能，他们发展的知识以及他们发现的黑洞的自旋、脉动和无毛特征，与寻找和解释天鹅X-1没有一点儿关系。如果天鹅X-1有确凿的黑洞信号，情形可能会不同，但事实上没有。

这些年轻人和其他跟他们一样的理论物理学家有时被人称为相对论学家，因为他们很多时间都在跟广义相对论的定律打交道。但真正投身寻找黑洞的理论家们（泽尔多维奇、帕津斯基、桑尼耶夫、里斯，等等）是全然不同的一群，他们被称为天体物理学家。对寻找来说，这些天体物理学家只需要懂很少一点广义相对论——只要能让他

们相信弯曲时空与此毫不相干，而牛顿的引力图景对模拟像天鹅X-1那样的天体已经足够了。但是，他们需要大量别的知识，那些一个天体物理学家所必须具备的东西。他们需要掌握广博的天文学专业知识，例如，关于双星系的，关于候选黑洞的伴星的结构、演化和光谱的，以及关于星际尘埃对星光的红移影响的——这是确定天鹅X-1距离的关键。他们需要懂的东西还多着呢：热气流，热气流碰撞形成的激波，湍流和无规则磁场引起的气体摩擦，磁力线的断裂和重新联结，热气体中X射线的形成，X射线在气体中的传播等许许多多东西。由于人类的极限，很少有人能掌握这么多东西并同时掌握复杂的弯曲时空的数学，所以研究者群体必然会有分工。你可能专攻黑洞的理论物理学，从广义相对论推演黑洞应有的性质；你也可能熟悉双星系和落在黑洞上的热气体以及这些气体产生的辐射；就是说，你要么是一个相对论学家，要么是一个天体物理学家。

也有人既想做相对论学家，也想做天体物理学家，但并不太成功。泽尔多维奇是卓有成就的天体物理学家，有时也能提出一些关于黑洞基础的新见解。我多少算一个有才能的相对论学家，曾尝试建立天鹅X-1的黑洞附近流动气体的广义相对论模型。但是，泽尔多维奇对广义相对论认识不深，而我也不太熟悉天文学的专业知识。横在两家间的障碍是巨大的。在我所认识的黄金年代的研究者中，只有诺维科夫和钱德拉塞卡，一脚坚实地扎根在天体物理学，另一脚扎根在相对论。

像贾柯尼那样的实验物理学家，设计并放飞过X射线探测器和卫星，也面临着相同的障碍。但有一点不同。在黑洞的找寻中，相对论学家是不需要的，而实验物理学家却是基本的。如果没有实验物理学家提供数据，那些把握了认识双星、气流和X射线传播工具的实测天文学家和天体物理学家们什么也做不成。实验物理学家常常在将数据交给天文学家和天体物理学家之前，自己试着去发现，关于气流和可能产生这些气流的黑洞，这些数据会告诉我们些什么。但没能获得多

大成功。天文学家和天体物理学家感谢他们提供了这些数据，然后以自己更复杂也更可靠的方式来解释它们。

天文学家和天体物理学家对实验物理学家的依赖，只是一个例子，为了成功寻找黑洞，还有许多重要的相互依赖关系。实际上，成功是6个不同群体相互依靠、协同努力的结果。每一个群体都扮演着重要的角色。相对论学家根据广义相对论的定律肯定了黑洞的存在。天体物理学家提出了寻找黑洞的方法并在几个关键步骤上给予了指导。实测天文学家确认了天鹅X-1的伴星，HDE 226868；用从它发出的周期性移动谱线来估计天鹅X-1的质量；并以大量其他观测来巩固他们的质量估计。实验物理学家创造的仪器和技术使X射线星的寻找成为可能，并在天鹅X-1的寻找中实现了。美国宇航局（NASA）的工程师和管理者们制造了把X射线探测器送入地球轨道的火箭和卫星。而且，同样重要的还有，美国的纳税人为火箭、飞船、X射线探测器和X射线望远镜，以及工程师、管理者和同他们一起工作的科学家们，提供了几亿美元的资助。

凭着这些令人难忘的团结协作，现在，90年代，我们几乎百分之百地相信，不仅在天鹅X-1，而且在我们星系的其他许多双星系中，都存在着黑洞。

[1] 《阁楼》杂志是与《花花公子》一样出名的美国男性杂志（它们都不是某些人所想象的那种“色情杂志”；在美国还有不少人说它们是教育性杂志呢），所以令作者一家的女性都不乐意。（上回作者与钱德拉塞卡赌赢了，也没敢要《花花公子》，读者大概还记得。）——译者

[2] 照片上的赌约写的是：“鉴于S·霍金对广义相对论和黑洞素有研究且需要一份保险，而K·索恩好冒险，不需要买保险，故以打赌定胜负。霍金以1年《阁楼》对索恩4年《私家侦探》，赌天鹅X-1不含质量大于钱德拉塞卡极限的黑洞。”下面是两人和证人签名，时间是1974年12月10日。霍金后来写的是：“认输，1990年6月”。——译者

[3] 见第7章。带电黑洞的电场平均地分布在旋转轴的周围，所以不可能产生集中的辐射束。

# 第9章　意外发现

天文学家先前没有料到，

现在被迫承认，

在星系的中心，

可能存在比太阳重100万倍的黑洞

## 射电星系

如果有谁在1962年（理论物理学家刚开始接受黑洞概念的那一年）断言，宇宙包含着比太阳重数百万或数十亿倍的巨大黑洞，天文学家一定会笑他。不过，天文学家不知道，他们从1939年起就已经在用无线电波观测这样的巨黑洞了。至少我们今天会强烈地这么猜测。

无线电波是与X射线相对的另一个极端。X射线是波长极短（典型波长比可见光短10　000倍）的电磁波（见序幕的图P.2）；无线电波也是电磁波，但波长很长，波峰到波峰的典型距离为几米，比可见光长百万倍。从波粒二象性说（卡片4.1），X射线与无线电波也处在两个相对的极端——电磁波有时像波，有时像粒子（光子），X射线的典型行为就像高能粒子（光子），X射线光子击中原子，从原子中打出电子（第8章），这样就很容易用盖革计数器来探测。无线电波几乎总是表现为电力和磁力的波，很容易用金属或天线来探测，因为电力的振荡能使电子上下振动，从而在天线上固定的无线电接收器中产生振荡信号。

宇宙无线电波（或射电波，来自地球外面的无线电波）是央斯基（Karl　Jansky）1932年偶然发现的，他那时是新泽西州霍尔姆德尔的贝尔电话实验室的无线电工程师。[1]刚从大学出来，央斯基就被派去识别干扰联通欧洲电话的噪音。那个时候，通过大西洋的电话是靠无线电传输的，所以央斯基做了一架由长长的金属管构成的特殊无线电天线，来寻找无线电静电干扰的来源（图9.1（a））。他很快发现，多数干扰来自雷雨，但雷雨过后还残留有微弱的嘶嘶噪音。到1935年，他已经确定了那噪音的来源，它很可能来自我们银河系的中心区域。当中心区域在头顶时，噪音较强，当它沉到地平线以下时，噪音会减弱，但不会完全消失。

这是一个令人惊异的发现，任何想过宇宙电波的人都会认为太阳是天空最强的无线电波源，就像它是最亮的光源一样。毕竟，太阳比银河系中大多数恒星离我们近10亿（$10^9$）倍，所以它的无线电波应该比来自其他恒星的强约$10^9\times10^9=10^{18}$倍。因为在我们的星系中只有$10^{12}$颗恒星，所以太阳应该比所有其他恒星加在一起还亮大概$10^{18}/10^{12}=10^6$（100万）倍。这样论证怎么会错呢？来自遥远的银河中心的无线电波怎么会比来自太阳的强那么多呢？

这当然是令人惊奇的谜，但回想起来，还有更令人惊奇的事，那就是天文学家对这个谜几乎一点儿也没留意。实际上，尽管贝尔电话公司极惹人注意，但央斯基的发现似乎只有两个天文学家表现了一点儿兴趣。它注定了会因天文学家的保守而被埋没，同当年钱德拉塞卡宣布没有重于1.4个太阳质量的白矮星的遭遇一样（第4章）。

这两位例外的热心人，来自哈佛大学天文系，一个是研究生格林斯坦（Jesse Greenstein），一个是讲师惠普尔（Fred Whipple）。他们在认真考虑了央斯基的发现后证明，如果当时流行的关于宇宙电波产生的思想是正确的，那么我们的银河系不可能产生央斯基所发现的那么强的电波。[2]尽管显然不可能，格林斯坦和惠普尔还是相信央斯基的发现。他们确信问题出在天体物理学理论，而不在央斯基。但理

论哪儿错了，一点线索也没有；另外，正如格林斯坦后来回忆的，“[30年代]我也没有碰到对这件事情感兴趣的天文学家，一个也没有。”[3]所以，他们也将注意力转向了别的地方。

1935年（大约茨维基提出中子星概念的时候，见第5章），央斯基已经完全认识了他的原始天线所能发现的银河系噪音。为认识更多的东西，他向贝尔电话实验室建议建造世界第一台射电望远镜；那是一个直径100英尺（30米）的巨大金属碗，它会像光学反射望远镜把光从镜片反射到目镜或摄影板上那样，将传来的电波反射到无线电天线和接收器。贝尔公司否决了这项建议，因为它不会带来好处。央斯基是个好雇员，也只得同意。他放弃了对天空的研究，在第二次世界大战临近的阴影中，将精力转向了短波长的无线电通讯。

职业科学家对央斯基的发现实在太没有兴趣了，在接下来的10年里才有第一个人来造射电望远镜，这人是雷伯（Grote Reber），一个古怪的单身汉，伊利诺斯惠顿的业余无线电接线员，呼号W9GFZ。[4]他从《大众天文学》杂志读到央斯基的射电噪音后，就开始研究它的细节。雷伯没受过多少科学教育，但那并不重要。重要的是他有良好的无线电工程训练和强烈的实践精神。凭他巨大的创造力和有限的积蓄，他靠自己的双手在母亲的后院设计制造了世界上第一台射电望远镜，镜面是直径30英尺（也就是9米）的盘子；他用它绘出了天空的射电图（图9.1（d））。从他的图中可以清楚地看到，射电源除了我们银河系的中心区外还有两个，后来被称为CygA和CasA——A代表“最亮的射电源”，Cyg和Cas分别代表“天鹅座”和“仙后座”（Cassiopeia）。40多年的研究最终证明，天鹅A和许多随后发现的射电源，很可能是由巨黑洞提供能源的。

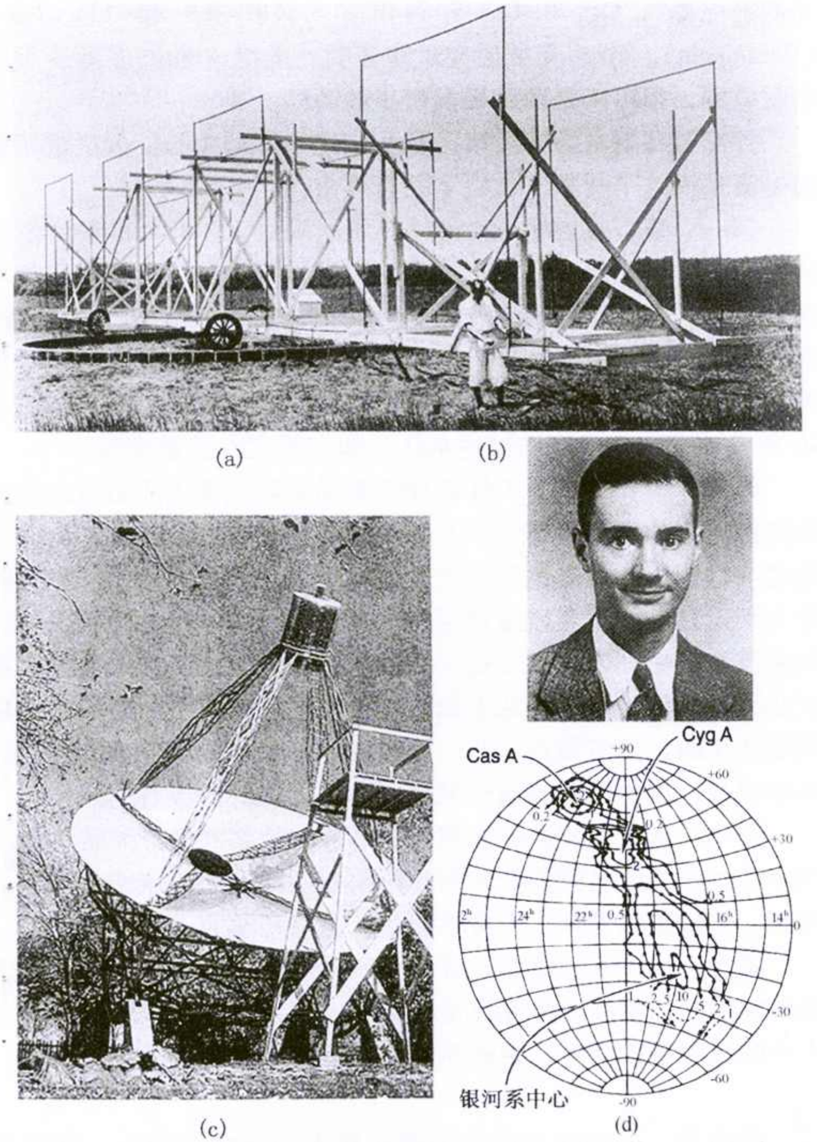



图9.1　（a）央斯基和他1932年发现来自银河的宇宙电波的天线。（b）G·雷伯，约1940年。（c）雷伯建在伊利诺斯惠顿他母亲后院的世界第一台射电望远镜。（d）雷伯用他的射电望远镜绘出的天空射电图。[照片（a）由贝尔电话实验室摄，美国物理学联合会（AIP）Emilio Segre图像档案馆提供；（b），（c），（d）由雷伯提供；（d）引自Reber（1944）。]

探测这些射电源的故事，是我们这一章的中心。我决定用整整一

章来讲这个故事，有以下几点理由：

第一，故事将说明一个与第8章所讲的大不相同的科学发现的模式。在第8章，泽尔多维奇和诺维科夫提出了具体的寻找黑洞的方法；实验物理学家、天文学家和天体物理学家实现了那个方法，而且成功了。在这一章里，雷伯在1939年就观测到巨黑洞了，那时还没人想到寻找它们，而越来越多的观测证据等了40年，才令天文学家们被迫承认，他们看到的就是黑洞。

第二，第8章讲了天体物理学家和相对论学家的力量；这一章要讲他们的局限。第8章发现的那类黑洞在人们寻找它们的四分之一世纪以前就被预告存在了，那是奥本海默-斯尼德黑洞：比太阳重几倍，由大质量恒星坍缩而成。本章的巨黑洞不一样，从来没有哪个理论家预言过它们的存在；它们比任何天文学家在天空见过的任何恒星重几千或几百万倍，所以不可能是那些恒星坍缩产生的。任何预言这些巨黑洞的理论家可能都会损害自己的科学荣誉。这些黑洞的偶然发现，才是真正的“发现”。

第三，这一章要讲的发现的故事，将比第8章更清楚地说明四个科学家群体之间复杂的相互影响和依赖关系，他们是相对论学家、天体物理学家、天文学家和实验物理学家。

第四，这一章的最后要证明，巨黑洞的自旋和转动能量在解释所观测到的电波中将发挥重要作用。相反，对第8章的那些不太大的黑洞性质而言，旋转是无关紧要的。

1940年，雷伯在完成他的第一次天空射电扫瞄后，认真地将他的望远镜、观测结果和射电图整理成一篇论文，寄给钱德拉塞卡，那时他在威斯康星日内瓦湖畔的芝加哥大学叶凯士天文台担任《天体物理学杂志》的编辑。钱德拉塞卡把雷伯不同寻常的稿子在叶凯士的天文学家中传阅。文稿令人困惑，而这位完全默默无闻的业余作者也令人

生疑。于是，几个天文学家驱车来到惠顿，看他的仪器。他们感动了。回去后，钱德拉塞卡同意论文发表。[5]

格林斯坦在哈佛读完研究生后，也成了叶凯士的一位天文学家，接下来的几年里，他多次来到惠顿，成了雷伯的好朋友。他说雷伯是“理想的美国发明家。假如他没对射电天文学发生兴趣，早就成百万富翁了”。[6]

格林斯坦对雷伯的研究满怀热情，几年后，他设法让他到芝加哥大学来。“学校不愿在射电天文学上花一分钱，”格林斯坦回忆说；但叶凯士天文台台长斯特鲁维（Otto Strave）同意给他一个研究职位，不过，雷伯的薪水和研究资助得从华盛顿来。然而，雷伯“是独来独往的怪人”，[7]格林斯坦说，他拒绝向当局详细报告如何使用造新望远镜的经费。于是，事情没办成。

这时候，第二次世界大战结束了，为战争做技术研究的科学家们开始寻找新的挑战。他们中间，有在战争中发明过跟踪敌人飞行器雷达的实验物理学家。雷达无非就是让无线电波从射电望远镜式的发射器发出，在飞机上反弹，然后回到发射器。所以让这些实验物理学家投身到射电天文学的新领域中来，是再理想不过了——有些人正渴望着来。巨大的技术挑战在等着他们，理性的回报也大有希望。在众多想大显身手的人中，三个小组立刻就占领了这片领域：英国约德雷尔邦克-曼彻斯特大学的洛弗尔（Bernard Lovell）小组，剑桥大学的赖尔（Martin Ryle）小组以及澳大利亚博塞（J.L.Pawsey）和波尔顿（John Bolton）的联合小组。在美国，没有值得注意的行动；雷伯差不多还是独自继续着他的射电天文学研究。

光学天文学家（用光来研究天空的天文学家，[1]那时就只有这一类天文学家）对实验物理学家的火热行动并不在意。要等到射电望远镜能准确测量射电源在天空的位置并能确定电波来自哪一个发光体时，他们才会发生兴趣。这要求望远镜的分辨率在雷伯达到的精度上

提高100倍，也就是被测射电源的位置、大小和形状的精度还应该提高100倍。

这个要求很高。光学望远镜，甚至人眼都很容易达到较高的分辨水平，因为它们依靠的（光）波具有很短的波长，小于$10^{-6}$米。反过来，耳朵不能很准确地区别声音所来的距离，因为声波的波长很长，大概1米左右。同样，无线电波因为具有米量级的波长，所以分辨率很低——除非望远镜远大于1米。雷伯的望远镜不是很大，分辨率当然也不太高。为了提高100倍的分辨率，需要大100倍，即约1公里的望远镜，或者需要用更短波长的无线电波，例如用几厘米的波来代替1米的波。

1949年，实验物理学家实现了那100倍，他们没有蛮干，而靠的是机智。通过简单而熟悉的例子，就能理解他们机智的想法。（这只是一个类比；说得太轻松了，不过能让我们对一般思想有个印象。）人仅凭两只眼睛，不用更多，就能看出我们周围的世界是三维的。左眼看到一点儿物体的左面，右眼看到一点儿物体的右面。把头转到一个面上，可以看到一点儿物体的顶面和底面；如果眼睛离远些（结果就像两台相机照三维照片，会有些夸张），我们也多少能看得更多一些。然而，我们的三维图像不会因为脸上长满了眼睛而有多大的改进。眼睛多了，我们看东西会明亮得多（也就是有更高的灵敏度），但在三维分辨水平上，我们并不能获得什么。

现在，1公里的巨大射电望远镜（图9.2左）多少有点儿像长满了眼睛的脸。望远镜是1公里大的碗，用能将电波反射和聚焦到线阵天线和接收器上的金属片覆盖。如果将各处的金属片拿走，只留下几片散布在碗上，就相当于将多余的眼睛从我们脸上拿走，只留下两只。这两种情形，都不会损失分辨率，只是极大地损失了灵敏度。实验天文学家最需要的是提高分辨率（他们想发现电波从哪儿来，波源是什么形状），而不是提高灵敏度（不需要看到更多更暗的射电源——至少目前不需要）。于是，他们只需要一个斑斑点点的碗，而不需要被

完全覆盖的碗。

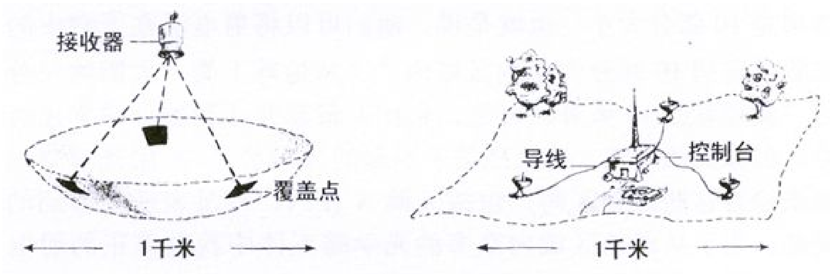



图9.2　射电干涉仪原理。左：为达到良好的角分辨率，可能有人想有一个巨大的，如1公里的射电望远镜。不过，电波反射碗只需要几点（黑点）用金属片覆盖并发生反射就够了。右：从那些点反射的电波不必都聚焦到一个天线和巨碗中心的接收器上；每一点可以聚焦到各自的天线和接收器，然后将各接收器的最终信号通过线路传到一个中心接收站，电波在接收站合成，与它在望远镜的接收器上合成一样。结果，这是一个具有相关而合成输出的小射电望远镜网，也就是一个射电干涉仪。

这个斑斑点点的碗，实际上是一个小射电望远镜的网络，通过线路联结到一个中心电波接收站（图9.2右）。每个小望远镜就像大碗上的一块金属片，每个小望远镜的射电信号通过线路传递，就像射电束在大碗的金属片上反射；合成来自各线路信号的中心接收站，也就像大碗上合成来自碗上各点的反射束的天线和接收器。这样的小望远镜和实验物理学家的中心工作站组成的网络，叫射电干涉仪，因为它背后的工作原理是干涉测量法，通过各小望远镜的输出结果的相互“干涉”（如何干涉，我们将在第10章卡片10.3中看到），中心接收器将合成天空的射电图或图像。

从40年代后期到50年代，进入60年代，那三个实验物理学家小组（约德雷尔邦克的、剑桥的和澳大利亚的）相互竞争着做更大和更灵巧的射电干涉仪，分辨率也前所未有地提高了。在光学天文学家中激起兴趣的第一个决定性的100倍的提高在1949年实现了。那时，波尔顿、斯坦莱（Golden Stanley）和澳大利亚小组的斯里（Bruce Slee）关于许多射电源的位置的误差区间是10弧分大小，也就是说，他们可以将射电源在天空中的位置确定到10弧分大小的区域内。

8（从地球上看，太阳的三分之一直径就是10弧分，可见，它比人眼靠光达到的分辨率还可怜，但靠无线电波，这已经是了不起的分辨率了。）用光学望远镜来检查这些误差区间，包括天鹅A在内，都没表现出特别的明亮；为了从误差区域内众多的光学暗天体中找到真正的射电源，电波的分辨率还需要提高。不过，这些误差区域中还是有三个与众不同的光亮天体：一个是古老超新星的残骸，另外两个是遥远的星系。

央斯基发现的从我们银河系发出的无线电波，已经令天体物理学家难以解释，现在要理解遥远星系怎么能发出那么强的无线电信号，就更困难了。天空中某些最亮的射电源可能是极遥远的天体，这令人难以置信（尽管后来证明真是这样）。于是，似乎可以打赌，每个误差区间里的无线电信号不是来自遥远的星系，而是来自区域内的某个光学暗淡然而距离很近的恒星。（打这个赌的人要输。）只有更高的分辨率能确定地告诉我们。实验物理学家在努力向前，几个光学天文学家开始产生了一点兴趣，在一边看着。

到1951年夏，剑桥的赖尔小组将分辨率又提高了10倍，赖尔的研究生史密斯（Graham　Smith）用它将天鹅A的位置确定到1弧分的误差区间——这个区间够小了，大概只含有100个左右的光学天体（用光看到的天体）。史密斯把他猜测的最佳位置和误差区间寄给帕萨迪纳卡内基研究所的著名光学天文学家巴德（就是17年前跟茨维基一起确认超新星并提出其中子星能源的那个巴德——见第5章）。卡内基研究所在威尔逊山上有一台那时世界上最大的2.5米（100英寸）光学望远镜；帕萨迪纳街上的加州理工学院刚在帕洛玛山上建好一台5米望远镜。研究所和学院的天文学家共用这些望远镜。巴德在他接下来计划的帕洛玛5米镜（图9-3（a））的观测中，拍摄了史密斯说的天鹅A所在天空的误差区域。（这个区域同大多数区域一样，以前没有用大光学望远镜检验过。）照片冲洗出来后，巴德简直不敢相信自己的眼睛：误差区内有一个从没见过的天体，仿佛是两个星系撞在

一起（图9.3（d）中间）。[9]（通过80年代红外望远镜的观测，现在我们知道那是光的错觉。天鹅A真是一个星系，不过有一尘埃带穿过它的表面，尘埃吸收了光，结果一个星系看起来就像两个撞在一起的星系。）中心星系加上射电源这样一个总系统，以后被称为射电星系。

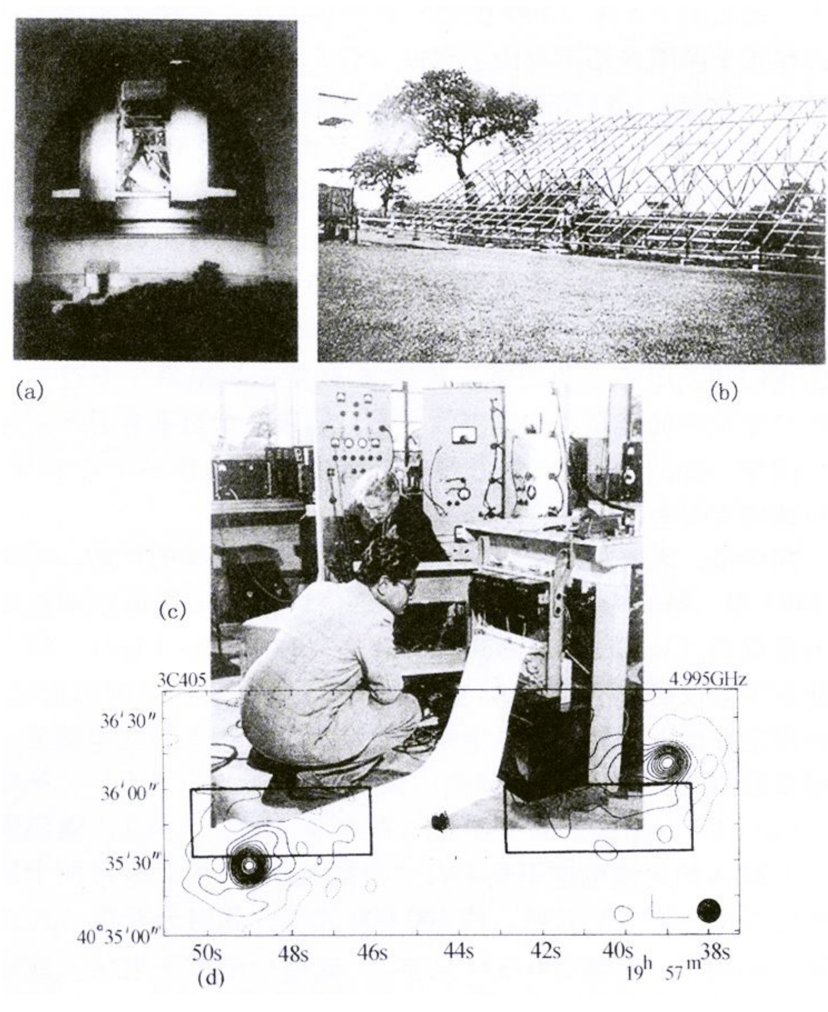



图9.3　发现天鹅A是一个遥远的射电星系：（a）5米光学望远镜，1951年巴德发现天鹅A关联着两个正在碰撞的星系。（b）约德雷尔邦克的射电干涉仪。1953年，詹尼森和古普塔用它证明，无线电波来自碰撞星系外的两片巨叶。干涉仪的两架天线（每一天线都是木架上的线

阵）也并排在图上。测量时，一架天线让卡车载着在地上跑圈儿，另一架留在原地。（c）詹尼森和古普塔在干涉仪控制室内检查无线电数据。（d）1953年观测所揭示的射电源的两叶（矩形）；同时，图中间还有巴德的“碰撞星系”的光学照片。图（d）还表现了高精度的无线电发射叶的等值线图（细实线），那是剑桥赖尔小组1969年得到的结果。[（a）由加利福尼亚理工学院帕洛玛天文台提供；（b），（c）由曼彻斯特大学努菲尔德射电天文台提供；（d）引自Mitton and Ryle（1969）、巴德和闵可夫斯基（1954）、Jennison and Das Gupta（1953）。]。

两年里，天文学家们相信，射电波是星系碰撞的产物。可是到1953年，新的惊奇出现了。那年，洛弗尔的约德雷尔邦克小组的詹尼森（R.C.Jennison）和古普塔（M.K.Das Gupta）用一种新式干涉仪研究了天鹅A。他们的干涉仪由两台望远镜组成，一台固定在地上，另一台让卡车载着绕圈子，为的是逐步覆盖一台想象的40平方公里望远镜的“大碗”上众多的“点”（见图9.2（b）左）。靠这个新干涉仪（图9.3（b），（c）），他们发现，天鹅A的无线电波不是来自“碰撞星系”，而是来自两个巨大的近似矩形的空间区域，约200 000光年，相对分开在“碰撞星系”的两头，距离200 000光年。[10]这两个电波发射区，或者叫“叶”，表现为图9.3（d）的两个矩形区。另外，巴德的“碰撞星系”的光学照片也在图的中间。在图中还可以看到叶状电波发射区更具体的情况，那是16年后更精密的干涉仪的成果；图中表现射电波亮度的细等值线，与地图上表现地形高度的等高线是一样的。等值线证实了1953年的结论：射电波来自“碰撞星系”两端的巨大气体叶。一个巨黑洞如何能为这两片巨叶提供能量，是本章后面的主题。

这些令人惊异的发现，在光学天文学家中间激起了长久而强烈的兴趣。格林斯坦不再是惟一认真的关注者了。

在格林斯坦个人看来，这些发现是最后的一棵稻草了。美国人没能在战后推进射电研究，现在他们只能站在一边儿看着这场自伽利略发明望远镜以来最伟大的天文学革命。革命的果实正在不列颠和澳大利亚成熟，而美国什么也没有。

格林斯坦这时是加州理工学院的教授，他从叶凯士来这儿，是为了围绕5米的新望远镜确立一个天文学计划。于是，他然去找院长杜布里奇（Lee Dubridge），要加州理工学院造一台射电望远镜，与5米望远镜一起探索遥远星系。杜布里奇在战时曾负责美国雷达设计，他有同感但也很谨慎。为说服杜布里奇行动起来，格林斯坦1954年1月5日和6日，在华盛顿组织了一次关于射电天文学未来的国际会议。11

在华盛顿，来自英国和澳大利亚的射电观测者们报告了他们惊人的发现，接着，格林斯坦提出了他的问题：一定要让美国继续成为射电天文学的荒漠吗？答案是显然的。

在国家科学基金会的大力支持下，美国的物理学家、工程师和天文学家们开始在西弗吉尼亚格林邦克建立国家射电天文台；杜布里奇也批准了格林斯坦关于建造最先进的加州理工学院射电干涉仪的报告，决定将它建在加利福尼亚约色米特国家公园西南的欧文斯河谷。由于学院里没人造过这种仪器，格林斯坦就把波尔顿从澳大利亚请来，做这个计划的先锋。

[1] 在本书中，“光”总是指人眼可见的那一类电磁波，也就是光学辐射。

# 类星体

到50年代后期，美国人也赶上来了。格林邦克的射电望远镜正在投入使用；在加州理工学院，马修斯（Tom Mathews），马尔特比（Per Eugen Maltby）和莫菲特（Alan Moffett）用欧文斯河谷的新射电干涉仪，与巴德、格林斯坦和其他人用帕洛玛5米光学望远镜一起，发现和研究了大量的射电星系。

1960年，这些工作又带来一个惊奇：加州理工学院的马修斯收到帕尔默（Henry Palmer）的来信。根据约德雷尔邦克的观测，一个叫3C48（剑桥的赖尔小组编制的第3版星表[1]中的第48个射电源）的射电源极小，不超过1弧秒的直径（太阳张角的1/10 000）。这么小的源是很新奇的事情。但是，帕尔默和约德雷尔邦克的同事们却不能很好地确定源的位置。马修斯在学院的新射电干涉仪上的工作非常出色，他将位置定到只有5弧秒大小的误差区域，把结果给了帕萨迪纳卡内基研究所的光学天文学家桑德奇（Allan Sandage）。桑德奇接着在5米光学望远镜的观测中，拍摄了马修斯误差区域中心的照片，他惊讶地发现，那不是一个星系，而是一个单独的蓝色亮点，像一颗恒星。“第二天晚上，我检查了它的光谱，那是我见过的最离奇的光谱，”他后来回忆说。谱线的波长根本不像恒星或地球产生的热气体的；也不像天文学家和物理学家以前遇到过的任何东西。从这个奇怪的天体，桑德奇看不出一点儿意思。

接下来的两年里，又有6个相同的天体以相同的方式被发现了，它们都跟3C48一样令人疑惑。加州和卡内基的所有光学天文学家都开始来为它们摄像，取光谱，力图认识它们的本质。答案本应是很显然的，但实际上不是，遇到阻碍了。这些奇异的天体那么像恒星，于是天文学家一直试图把它们解释成我们以前没有见过的银河系中的某种恒星，但这些解释太牵强，不会有人相信。

打破障碍的是32岁的荷兰天文学家施米特（Maarten Schmidt），[12]那时他刚应聘来到加州理工学院。几个月来，他都在想办法去认识他得到的3C273的光谱，那也是一个奇异天体。1963年2月5日，他坐在学院办公室里仔细地为他正在写的一篇文章画光谱，答案突然降临了。光谱中4条最亮的线是氢原子气体产生的标准的“巴尔末线”——这是所有光谱线中最有名的，是大学生在量子力学课里学的第一类谱线。不过，这4条线并没有通常的波长，每一条都红移了16%。所以，3C273一定含有大量氢气，并且以16%的光速离开地球——比任何天文学家见过的任何恒星的速度大得多。

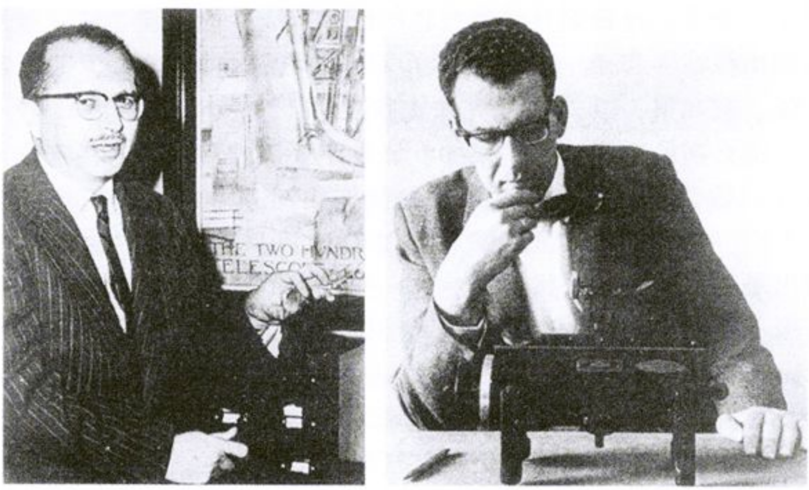
施米特冲出办公室，跑去找格林斯坦，激动地向他讲了自己的发现。格林斯坦回到办公室，把他的3C48光谱拿出来，盯着看了一会儿，没看到有任何红移的巴尔末线；但由镁、氧、氖发出的谱线在等着他，它们红移了37%。看来，3C48至少部分含有镁、氧和氖，以37%的光速离开地球。[13]

这么高的速度从哪儿来的？如果照普遍的想法，这些奇异天体（以后它们被称为类星体）是我们银河系中的某种恒星，那么它们一定是被巨大的力量从某处（也许从银河系中的核）喷射出来的，这太难以置信了。进一步检查类星体光谱会看到，这是极不可能的。格林斯坦和施米特认为（对的），惟一合理的解释是，这些类星体在我们宇宙很远的地方，由于宇宙膨胀的结果，它们在以极高的速度离开地球。

想想，宇宙膨胀就像正在吹气的气球的膨胀。假如有许多蚂蚁站在气球表面，那么每一只蚂蚁都会看到其他所有的蚂蚁在离它而去，这是气球膨胀的结果。离它越远的蚂蚁，离开它的速度也越快。同样，由于宇宙膨胀的结果，离地球越远的天体，我们在地球上看它离开的速度也越快。换句话说，天体的速度正比于它的距离。这样，施米特和格林斯坦能从3C273和3C48的速度推测它们的距离，分别是20亿光年和45亿光年。

这些距离太大了，几乎是有史以来的最大距离。这意味着，3C273和3C48为了达到在5米望远镜上显示的亮度，必然要辐射出大量的能量：比我们见过的最亮的星系的能量还高100倍。

3C273的确很亮，自1895年以来，它与它附近的其他天体一起，已被最普通大小的望远镜拍过2000多次了。在听说施米特的发现后，德克萨斯大学的史密斯（Harlan Smith）仔细检查了这些多数珍藏在哈佛档案馆里的照片，发现3C273在过去70年里亮度在波动。在短短1个月的时间内它发出的光发生过很大变化。[14]这意味着，大部分来自3C273的光必然是从一个比光在1个月内所经过的行程小，也就是从小于1“光月”的区域内发出的。（假如区域太大，当然就不会有任何以小于或等于光速运行的力量能使发出的气体同时在1个月内变亮和变暗。）



左：格林斯坦和帕洛玛5米光学望远镜的图，约1955年。右：施米特和他用来测量5米望远镜光谱的仪器，约1963年。[加利福尼亚理工学院档案馆提供。]

这些意思是极难让人相信的。3C273，这个奇异的类星体，比宇宙中最亮的星系还亮100倍。星系的光是从100 000光年大小的区域中发出的，而3C273的发光区域却只有1光月大小，直径至少比星系小

100万倍，体积小$10^{18}$倍。它的光必然来自一个由一台巨大功率的发动机加热的大质量气状天体。发动机很可能是一个巨大黑洞，不过还不能完全确信，15年过后才出现有力的证据。

如果说，解释来自我们银河系的无线电波难，解释来自遥远射电星系的无线电波更难，那么，解释来自那些超远类星体的无线电波，就难上加难了。

困难原来是一个思想障碍。格林斯坦、惠普尔和三四十年代所有天文学家一样，都认定宇宙电波跟恒星的光一样，是从原子、分子和电子的不规则热运动产生的，那时的天文学家不能想象还有别的自然途径能产生所看到的电波，尽管他们的计算已经确凿地证明原来的途径是行不通的。

然而，其他途径自20世纪初就已为物理学家所熟悉了：当高速运动的电子遇到磁场时，会因磁力作用而绕磁力线作螺旋运动（图9.4），在螺旋中发出电磁辐射。40年代，物理学家开始将这种辐射称为同步辐射，因为那时正在建造的所谓“同步”粒子加速器里的螺旋电子就产生这种辐射。值得注意的是，尽管物理学家对同步加速器表现了极大的兴趣，天文学家却毫不在意，头脑里的石头阻碍了他们的思想。

1950年，芝加哥的凯本海尔（Karl Otto Kiepenheuer）和莫斯科的金兹堡（就是为苏联氢弹发明LiD燃料，后来又发现黑洞无毛的第一丝线索的那个金兹堡[2]）打碎了那块石头。他们在阿尔文（Hans Alfvén）和赫洛森（Nicolai Herlofson）的思想基础上提出（对的），央斯基的来自我们银河的无线电波是螺旋地绕着充满在星际空间的磁力线近光速运动的电子产生的同步辐射（图9.4）。[15]

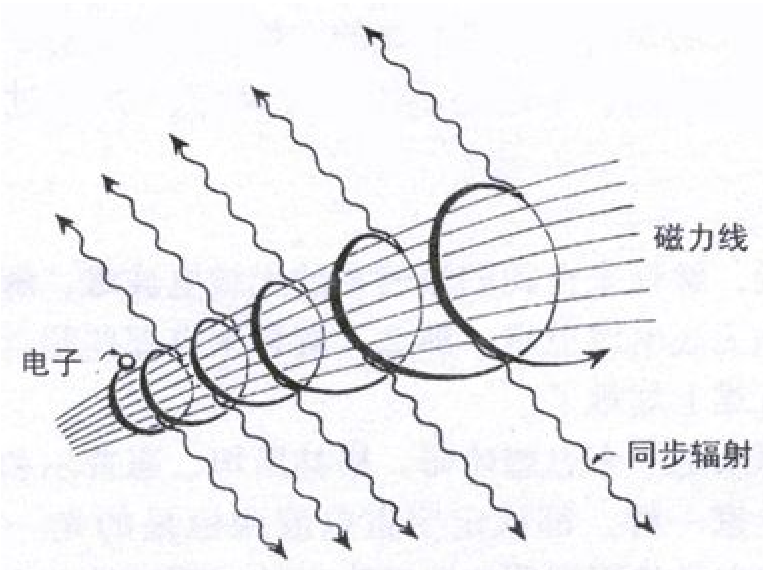



图9.4　宇宙的无线电波是在磁场中螺旋式近光速运动的电子产生的。磁场迫使电子螺旋而不是直线地运动，电子的螺旋运动产生无线电波。

几年后，发现了射电星系巨大的电波发射叶和类星体，人们自然（也是正确的）认为，它们的无线电波也是绕磁力线的螺旋电子产生的。根据螺旋运动和观测到的无线电波以及相关的物理学定律，圣地亚哥加利福尼亚大学的布尔比奇计算了射电叶的磁场和高速电子所应具有的能量。他的结果令人惊愕：在最极端情况下，电波发射叶必须具有的磁能和高速（动）能，相当于1000万（$10^7$）个太阳质量以100%的效率完全转化而来的纯粹能量。[16]

类星体和射电星系的能量要求那么大，1963年，天体物理学家们被迫去检验所有可以想象的能源。

化学能（汽油、油、煤或炸药的燃烧），这种人类文明的基本能源形式，显然是不够的。质量转化为能量的化学效率只有一亿分之一（$10^{-8}$）。为了向类星体发射电波的气体提供能量，需要$10^8\times10^7=10^{15}$个太阳质量的化学燃料——整个银河系所有燃料的100 000倍。这看来是完全不合理的。

核能，作为氢弹和太阳光热的基础，似乎是惟一能充当类星体能源的。核燃料的质能转化效率约为1%，所以，一个类星体需要$10^2 \times 10^7 = 10^9$（10亿）个太阳质量的核燃料来为它的电波发射叶提供能源。不过，只有当核燃料完全燃烧，而且能量完全转化为磁场和高速运动电子的能量时，10亿个太阳质量才够。完全的燃烧和完全的转化似乎是很不可能的。即使靠精密设计的机器，人类对燃烧能量的利用也很少超过几个百分点。而大自然没有什么周密的设计，很可能做得更差。于是，100亿或者1000亿个太阳质量的核燃料似乎更合理。这个量比一个巨大星系的质量小，但也不是特别小，而且大自然如何能够实现燃烧的核能向磁场能和动能转化，我们还不清楚。因此，核燃料是可能的，但可能性不是很大。

物质与反物质的湮灭[3]能实现100%的质能转化，所以1 000万个太阳质量的反物质与1 000万个太阳质量的物质发生湮灭，就能满足一个类星体的能量需要。然而，我们的宇宙中没有任何反物质存在的证据，只不过有一点在粒子加速器上人工产生的，和一点在物质粒子碰撞中自然产生的。而且，即使在类星体中有那么多物质和反物质湮灭，湮灭的能量将进入高能$\gamma$射线，而不会成为磁能和电子的动能。因此，物质-反物质湮灭似乎是很难令人满意的类星体供能方式。

还有一种可能：引力。正常恒星形成中子星或黑洞的坍缩，想来可能将10%的恒星质量转化为磁能和动能——虽然如何转换还不清楚。如果它真能这么做，那么$10 \times 10^7 = 10^8$（1亿）颗正常恒星的坍缩就可能满足1个类星体的能量，而像假想的比太阳重1亿倍的超大质量恒星，一颗就够了。[正确的想法是，这样一颗超大质量恒星的坍缩形成的巨大黑洞，本身可能就是为类星体提供动力的发动机。不过，在1963年还没人这么想过。那时，黑洞还没人理解；惠勒还没起“黑洞”的名字（第6章），萨尔皮特和泽尔多维奇还没认识到落向黑洞的气体可以高效地加热和辐射（第8章）；彭罗斯也还没发现黑洞可以将29%的质量作为旋转能贮存和释放（第7章）。黑洞研究的

黄金年代还没有来临。]

形成黑洞的恒星的坍缩可能为类星体提供能量，这个思想根本背离了传统。这是历史上第一次，天文学家和天体物理学家感到需要求助于广义相对论的效应来解释他们看到的天体。以前，相对论学家生活在一个世界，天文学家和天体物理学家生活在另一个世界，两家几乎没有交流。他们的偏见就要结束了。

为培养相对论学家与天文学家和天体物理学家之间的对话，促进类星体研究的进步，1963年12月12～18日，300名科学家在德克萨斯的达拉斯举行了第一届德克萨斯相对论天体物理学会议。[17]在一次餐后讲话中，康奈尔大学的戈尔德（Thomas Gold）描述了当时的情形，不过有点儿言不由衷："[类星体的神秘]令我们想到，相对论学者和他们精致的工作不仅是华丽的文化装点，也可能对科学真有些用处！现在，人人都高兴了：相对论学者们感到有人在感谢他们，成了他们几乎还不知道其存在的领域的专家；天体物理学家们也因为融和了另一个学科——广义相对论而扩展了自己的领地、自己的王国。这真是令人高兴的，让我们希望它是对的。如果到头来我们又和相对论学者们分开了，那将多令人羞愧。"

会议报告几乎从上午8:30持续到下午6:00，一个小时午餐；然后从下午6:00到凌晨2:00，进行非正式讨论和辩论。其中有一个10分钟的短报告，是年轻的新西兰数学家克尔作的，与会者都不认识他。克尔只是讲了他的一个爱因斯坦场方程的解——10年后会发现，这个解描述了旋转黑洞的所有性质，包括旋转能量的贮存和释放（第7，11章）；我们在下面将看到，这个解最终会成为解释类星体能量的基础。然而，在1963年，对大多数科学家来说，克尔的解似乎只是数学珍玩，甚至没人知道它描述了黑洞——尽管克尔猜想它也许能为旋转恒星的坍缩带来一点认识。

天文学家和天体物理学家来达拉斯是为了讨论类星体，对克尔神

秘的数学题目没有一丝兴趣。所以，当克尔上台讲话时，好多人溜出演讲厅到走廊上去谈他们喜欢的类星体理论；其他的更不礼貌的人，仍坐在大厅里小声讨论。留下的人许多都在打瞌睡，后半夜的科学讨论影响了睡眠，但靠这会儿是补不回来的。只有几个相对论学家在聚精会神地听。

这种场面，令帕帕皮特罗（Achilles Papapetro）忍不住了，他是世界有名的大相对论专家。等克尔一讲完，他就要求发言。他站起来，深有感触地解释了克尔功绩的重要性。他，帕帕皮特罗本人曾花30年时间寻找这样一个爱因斯坦方程的解，但同其他许多相对论学者一样，失败了。天文学家和天体物理学家礼貌地点点头；接下来，另一位报告者又来大讲类星体理论，他们又重新集中注意力，会议继续进行。[18]

60年代为射电源研究找到了一个转折点。以前，从事这项研究的人完全是实测天文学家——也就是光学天文学家和从事射电观测实验的物理学家，现在都团结到一个天文学群体中来了，叫射电天文学家。相反，理论天体物理学家没做什么事情，因为射电观测还没有细到能让他们很好地进行理论化的地步。他们惟一的贡献在于，认识了无线电波是由绕磁力线螺旋式高速运动的电子产生的，计算了它需要多大的磁能和动能。

60年代，随着射电望远镜分辨率持续提高和光学望远镜开始揭示射电源的新特征（例如，类星体小小的发光核），不断增长的信息源成了天体物理学家头脑的营养素。他们根据丰富的信息提出了许多解释射电星系和类星体的具体模型，而这些模型也一个个地被不断积累的观测数据所否定了。说到底，这就是科学的一贯作风！

关键的一点是，射电天文学家发现，无线电波不仅从处在射电星系两端的两叶，而且还从中央星系本身的核发出来。1971年，剑桥席艾玛的一个新来的学生里斯由此想到一个为两叶提供能源的新办法。

也许，星系所有的无线电波都来自星系核中的一台发动机，也许这台发动机在直接为中心的电波发射电子，为磁场提供能量，也许它还向两叶输送能量，激发那儿的电子和磁场；也许，这台在射电星系核心的发动机正是为类星体提供能量的那种（不管它可能是什么）。[19]

里斯原来怀疑，从星系核携能量到两叶的流由超低频的电磁波组成，但理论计算很快说明，这样的电磁流无论如何也不能穿过星系的星际气体。

里斯不太正确的想法引来了正确的思想，这倒是常有的事情。剑桥的朗盖尔（Malcolm Longair）、赖尔和谢维尔（Peter Scheuer）采纳了他的想法，做了一点简单的修改：他们留下里斯的流，但让它成为热的磁化气流，而不是电磁波束。[20]里斯立即同意这类气体喷流能实现那个过程，然后同他的学生布兰福德（Roger Blandford）一起计算了喷流应有的性质。

几年后，英国、荷兰和美国的新射电干涉仪辉煌地证实，发射电波的两叶是通过来自星系中央发动机的气体喷流获得能量的——其中最值得注意的是美国在新墨西哥州圣·奥古斯丁平原的VLA（甚大天线阵，图9.5）。这组干涉仪看到了喷流，而且正好具有所预言的性质。它们从星系中心看到两叶，甚至看到涌进来的气体在两叶慢慢停下来。

与四五十年代的射电干涉仪（图9.2）一样，VLA也采用“点盖碗”技术，不过它的碗大得多，点也多得多（联结着更多的射电望远镜）。它的分辨率达到了1弧秒，差不多与世界上最好的光学望远镜一样好——这是40年前人们对央斯基和雷伯的原始仪器所期待的巨大进步。但进步并没有就此停下。80年代初，由大陆或地球相对两端的射电望远镜组成的甚长基线干涉仪（VLBIs）得到了比光学望远镜的分辨率高1000倍的射电星系核和类星体的图片。（VLBI的各望远镜的结果记录在磁带上，并以一个原子钟为它们标记时间，然后将来自

所有望远镜的磁带输入计算机，在机上相互“干涉”而成图。）

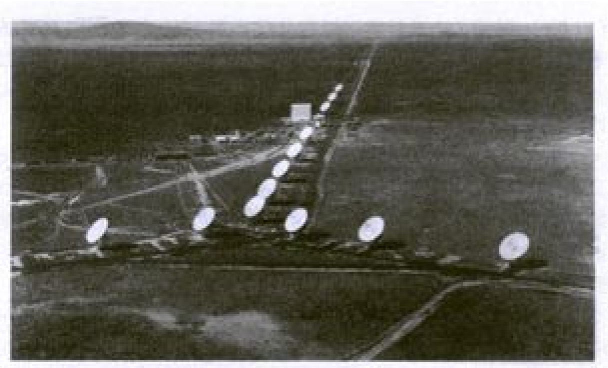
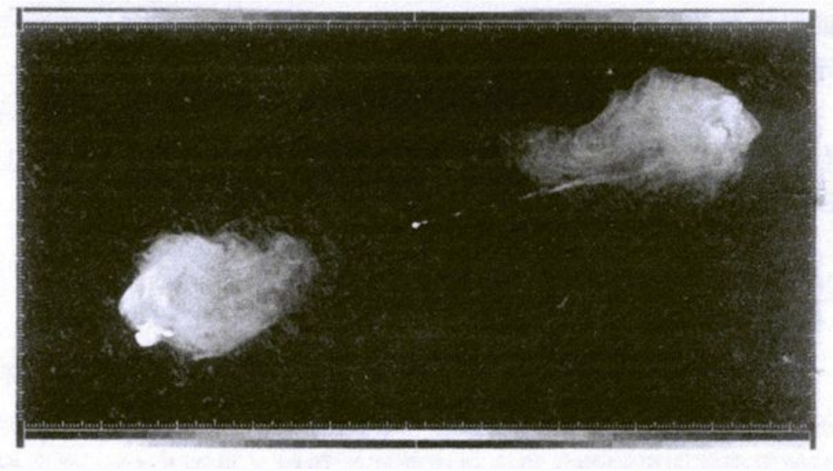


图9.5　上：新墨西哥州圣·奥古斯丁平原上的VLA射电干涉仪。下：珀莱（R.A.Perley），德雷耶（J.W.Dreyer）和科万（J.J.Cowan）用VLA拍摄的射电星系天鹅A的射电照片。涌入右射电叶的喷流很清楚；左叶的喷流暗一些。与1944年雷伯那幅没能反映两叶的对应照片（图9.1（d））和1953年詹尼森、古普塔那幅只揭示了两叶存在（图9.3（d）的矩形）的射电图以及1969年赖尔的图（9.3（d））比较，可以看到这幅射电图片在分辨率上大大提高了。[两图均由NRAO/AUI提供。]



80年代初的VLBI照片说明，喷流最深延伸到了星系核或类星体内几光年——就是在这个区域内，某些类星体（如3C273）藏着一个大小不足1光月的非常明亮的发光体。中央发动机大概也在发光体内，

不仅为它提供能量，也激发了涌向射电叶的喷流。

喷流还泄露了中央发动机本质的另一点线索。有的喷流在100万光年甚至更长的距离内是绝对直的。如果这些喷流的源在转动，那么像洒水车上旋转的水龙头一样，它会产生弯曲的喷流。所以，我们看到的直线喷流意味着，中央发动机在很长一段时间里是在完全相同方向上点燃喷流的。多长呢？由于喷流气体不能比光运动更快，而有些喷流比百万光年还长，所以点火方向必须稳定百万年以上。为达到这样的稳定性，发动机的喷流“龙头”必须固定在超稳定的天体——某种永久的陀螺仪上。（回想一下，陀螺是一种快速旋转的物体，它能长时间地将旋转轴保持在一个稳定不变的方向。这样的陀螺仪是飞机和导弹惯性导航系统的关键部件。）

到80年代初，在已经提出的10多个中央发动机解释中，只有一个需要永久的超级陀螺仪，它的大小不超过1光月，能产生强大的喷流。那是一个巨大的旋转的黑洞。

[1] 星表是天文学家用以记载各种天体参数的编目表，大概可以追溯到公元前4世纪中国石申的《星经》。星表种类很多，内容和用途不同。在一般读物中常看到的有《星云星团新总表》（NGC）、《梅西耶星表》（M）等。前面看到的HDE为哈佛大学天文台出版的光谱型星表；SS为“特殊星”，列在Hα发射线星表中；这里说的属于射电源表（《剑桥第三射电源表》）。——译者

[2] 见图7.3。金兹堡不仅以这些发现出名，他还有许多别的发现：与朗道发展超导体的“金兹堡-朗道理论”（为了解释为什么某些金属在很冷的情况下会完全失去对电流的阻力而提出的理论）。金兹堡是世界上几个真正的“文艺复兴式的物理学家”之一，也就是说，他几乎对理论物理学的所有分支都有重要贡献。

[3] 有关情况请看词汇表中“反物质”条以及第5章。

# 巨黑洞

巨黑洞可能激发类星体和射电星系的想法，是萨尔皮特和泽尔多维奇1964年（也就是黄金年代开始那年）提出来的。[21]他们曾发现落向黑洞的气流会碰撞而产生辐射（见图8.4），这个想法显然是那发现的一个结果。

关于气流向黑洞下落的更完整和实在的描述，是剑桥的英国天文学家林登-贝尔（Donald Lynden-Bell）1969年提出的。[22]他令人信服地证明了，气流碰撞后将结合在一起，在离心力作用下围绕黑洞螺旋式下落，在旋转中形成一个盘状物，就像围绕土星的环——林登-贝尔称它为吸积盘，因为气体是被吸到黑洞上去的。（图8.7右图是艺术家心目中的一个吸积盘，围绕着天鹅X-1中的一个小黑洞。）在吸积盘中，相邻气流会相互摩擦，强烈的摩擦会将盘加热到很高的温度。

80年代，天体物理学家认识到，3C273中心那个1光月左右大的明亮发光天体可能就是林登-贝尔说的那种摩擦生热的吸积盘。

我们通常认为，摩擦是很不起眼的热源。想想靠两根棍儿的摩擦来点火的可怜童子军吧！不过，童子军的肌肉力量太小了，而吸积盘的摩擦靠的是引力的能量。由于引力能很大，远大于核能，所以摩擦很容易把盘加热，使它比大多数明亮星系还亮100倍。

黑洞怎么像陀螺呢？1975年，耶鲁大学的巴丁（James Bardeen）和彼德森（Jacobus Petterson）找到了答案：[23]黑洞如果快速旋转，就完全像一只陀螺。它的旋转方向总是严格固定不变的，旋转在黑洞周围产生的空间旋涡（图7.7）也总是指着相同的方向。巴丁和彼德森通过数学计算证明，黑洞附近的空间旋涡一定把握着吸积盘的内部，使它严格保持在黑洞的赤道平面——不论盘的指向与黑洞方向相差多远，都是如此（图9.6）。来自星际空间的气体被盘的遥

远外部捕获时，可能会改变那里盘的指向，但由于黑洞的陀螺行为，它不可能改变黑洞附近的吸积盘方向。在黑洞附近，吸积盘总在赤道面上。

没有爱因斯坦场方程的克尔解，天体物理学家就不会认识这种陀螺行为，也不可能解释类星体。有了克尔解，他们在70年代中期就得到了一个清晰而优美的解释。黑洞不再是“空间的一个洞”，它作为一个动态物体的概念，第一次成为解释天文学家观测现象的重要角色。

巨黑洞附近的空间旋涡有多大？换句话说，巨黑洞旋转有多快？巴丁也导出了答案。他从数学上证明，从吸积盘落进黑洞的气体会逐渐使黑洞越转越快。黑洞在吞没了足够的螺旋气体而使自身质量加倍时，将以近乎最大可能的速率旋转——离心力的作用不允许超过这个速率（第7章）。[24]因此，巨黑洞通常应该有一个最大旋转速度。

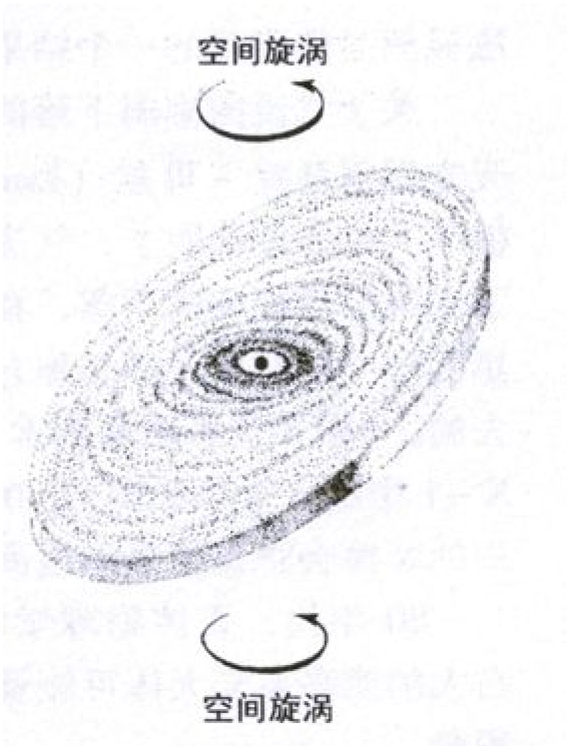



图9.6　旋转黑洞产生周围的空间旋涡，旋涡使吸积盘内部保持在黑洞的赤道面上。

黑洞和它的吸积盘怎么会产生两股在相对方向上的喷流呢？这太容易了。70年代中期，剑桥大学的布兰福德、里斯和林登-贝尔就知道，有四种可能产生喷流的途径，每一种都能做到这一点。

第一种：布兰福德和里斯认识到，[25]吸积盘周围可能是冷气体云（图9.7（a））。吹开吸积盘上下表面的风（类似于吹开太阳表面的风）可以在冷气云中生成热气泡。然后，热气体在冷云的上下表面钻孔，并从孔洞流出去。从这些孔洞流出来的热气体，像从花园里的浇水龙头流出的水一样，形成细细的喷流。喷流的方向依赖于孔洞的位置。假如冷气云也以黑洞的轴旋转，那么最可能的位置就是沿着转轴，也就是垂直于吸积盘的内部——这些位置的孔洞产生的喷流，就将固定在黑洞的陀螺旋转的方向上。

第二种：因为吸积盘很热，所以内部压力很高；而这个压力可能使盘膨胀，变得很厚（图9.7（b））。林登-贝尔指出，[26]在这种情况下，吸积盘气体的轨道运动的离心力，将在盘的顶部和底部面上生成旋涡式的漏斗，这些漏斗很像浴缸里的水从排水孔螺旋流出时形成的旋涡。黑洞像排出孔，而盘的气体就像水。由于气体的摩擦，旋涡状漏斗的表面也会很热，形成吹散自己的风，漏斗则将风汇聚成喷流。喷流方向与漏斗的一样，而漏斗牢牢地固定在黑洞陀螺的旋转轴上。

第三种：布兰福德发现，[27]处在吸积盘上和延伸到盘外的磁力线都会因盘的转动而被迫一圈圈地旋转（图9.7（c））。旋转的磁力线表现出向外和向上（或者向外和向下）的螺旋形状。电力将把热气体（等离子体）束缚在力线上；等离子体可以沿着力线滑动，但不能穿越。力线旋转时，离心力会将等离子体沿着力线抛出去，形成两股磁化喷流，一股向上，一股向下。这样，喷流仍然还是在黑洞的旋转方向上。

第四种喷流产生方式比其他几种更有意思，需要多说几句。在这种方式里，黑洞像图9.7（d）那样被磁力线穿过，它旋转时，也拖着磁力线转圈，使它们向上、下抛出等离子体，同第三种方法一样，形成两股喷流。喷流沿黑洞转轴射出来，所以它们的方向与黑洞陀螺是一样的。这种方式是布兰福德刚获剑桥博士学位后与另一个研究生茨纳耶克（Roman　Znajek）想到的，于是被称为布兰福德-茨纳耶克过程。[28]

布兰福德-茨纳耶克过程特别有意思的地方在于，流入喷流的能量来自黑洞巨大的旋转能。（这应该是显然的，因为正是黑洞的旋转引起空间旋涡，正是空间旋涡引起磁力线旋转，正是磁力线旋转将等离子体抛出去。）

在布兰福德-茨纳耶克过程中，黑洞的视界怎么可能被磁力线穿过呢？这样的力线本应是一种“毛”，可以转化为电磁辐射而消失，所

以，根据普赖斯定理（第7章），它们是必然会辐射掉的。事实上，普赖斯定理只有在黑洞远离其他物体而单独存在时才是正确的。我们现在讨论的黑洞并不孤立，在它周围还有吸积盘。假如那些磁力线突然脱离黑洞，那么从它北半球出去的和从它南半球出去的线实际上是互为延伸的同一条线，它们脱离黑洞的惟一途径是穿过吸积盘外部的热气体。但热气体不会让它们过去，而是将它们挤进吸积盘内部的空间区域。因为这个区域大部分被黑洞占据着，所以被困的磁力线穿过了黑洞。

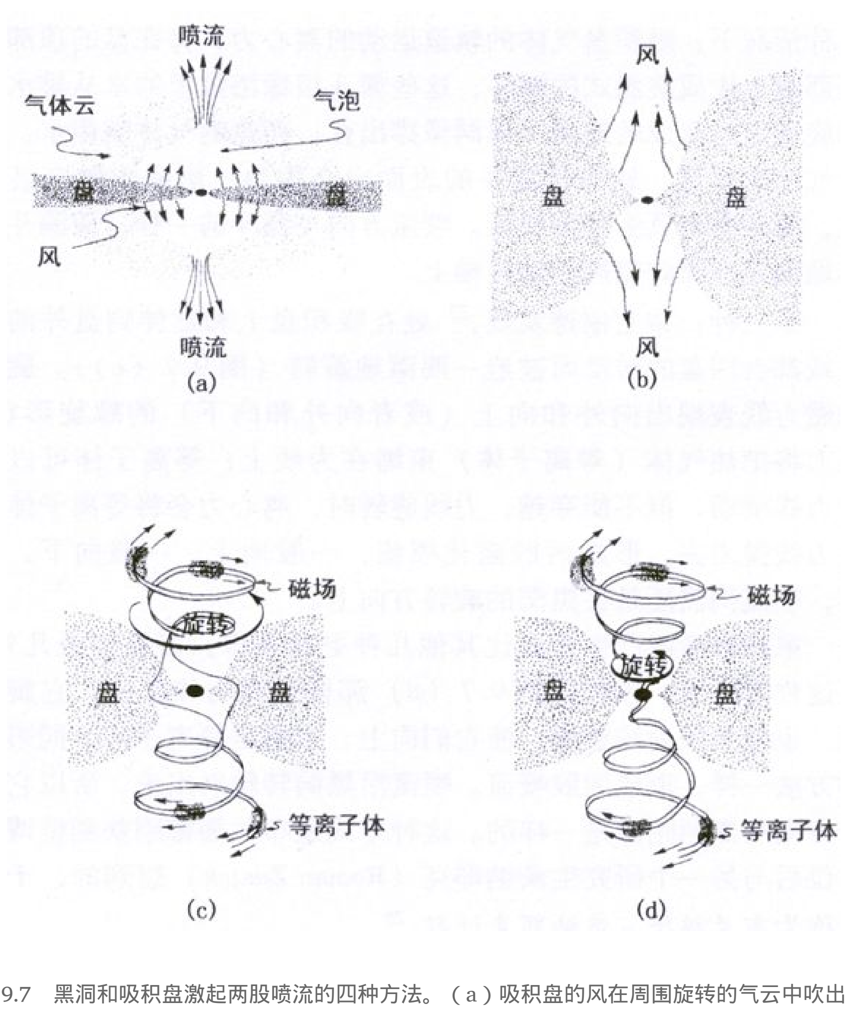



图9.7　黑洞和吸积盘激起两股喷流的四种方法。（a）吸积盘的风在周围旋转的气云中吹出一

个气泡，气泡的热气体沿旋转轴在气云中钻孔，喷流从孔洞中射出。（b）吸积盘因内部巨大的热压力而膨胀，膨胀后的旋转盘形成两个漏斗，吸积盘的风经过漏斗而成为喷流。（c）束缚在吸积盘上的磁力线被盘的旋转带动，磁力线旋转时，将等离子体向上下抛出去，等离子体沿力线滑动而形成两股磁化喷流。（d）穿过黑洞的磁力线因黑洞空间的旋涡而被迫旋转，它们在旋转时向上下抛出的等离子体形成两股磁化喷流。

这些磁力线从哪儿来？来自吸积盘。宇宙中所有气体都是磁化的，或至少有一点磁化，吸积盘的气体也不例外。[1]吸积盘气体一点点落入黑洞时，也将磁力线带进去了。在接近黑洞时，每一点气体从磁力线“滑”下去，穿过视界，而将力线留在视界外面，像图9.7（d）画的那样穿过视界。现在，这些被周围的吸积盘严格束缚着的穿过视界的磁力线，就可以照布兰福德-茨纳耶克过程提取黑洞的旋转能了。

所有这四种产生喷流的方式（气体云中的孔洞，漏斗里的风，吸积盘内旋涡式的磁力线和布兰福德-茨纳耶克过程），也许在类星体、在射电星系以及在某些其他类型星系的特殊核心（我们称这些核心是活动星系核），都不同程度地发挥着作用。

如果说类星体和射电星系的能源都来自相同类型的黑洞发动机，[29]它们为什么又显得那么不同呢？为什么类星体的光来自大小约1光月左右的强烈发光的恒星类天体，而射电星系的光来自大小约100 000光年的银河系那样的恒星集合？

几乎可以肯定，类星体与射电星系没有多大差别。类星体的中央发动机周围也存在着一个100 000光年的恒星系。不过，在类星体中，中央黑洞以特别高的效率通过吸积气体而增加燃料（图9.8），相应地，吸积盘内的摩擦热也很高。这么巨大的热量使吸积盘比周围星系的所有恒星加起来还亮几百到几千倍。天文学家看到了光亮的吸积盘，却没看到星系的恒星，所以这个天体看起来是“quasi-stellar”（也就是，像恒星那样的，像一个小小的强光点）[2]，而不像一个星系。

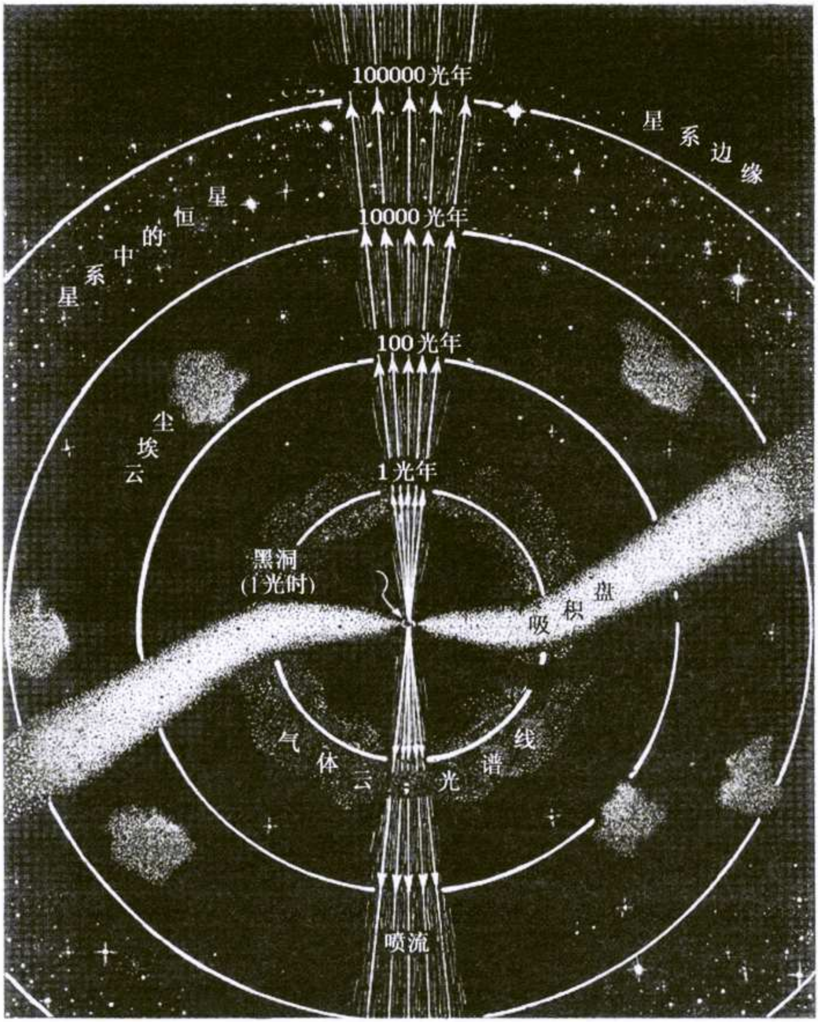



图9.8　我们现在所能理解的类星体和射电星系的结构。这个以所有观测数据为基础的具体模型是加州理工学院的芬尼（Phinney，E.Sterl）和其他人提出的。

吸积盘最深处很热，发X射线；靠外一点儿，吸积盘冷一些，发紫外辐射；再外就更冷，发光学辐射（可见光）；而在最外的区域，就冷却到只能发红外辐射了。发光区域的正常大小是1光年左右，不过在有些情况（如3C273）下，它可能只有1光月或更小，因此也可能在1个月那么短的周期内改变光度。从最内的区域流出的大量X射线

和紫外线落到离吸积盘几光年远的气体云上，将它们加热；正是这些被加热的云发出的光谱线，让我们第一次发现了类星体。有些（但不是所有的）类星体会出现吹散吸积盘的磁化风，这些风很强，足以产生射电的喷流。

与类星体相比，射电星系中央的吸积盘大概更安静一些，安静的意思是吸积盘内的摩擦很小，于是热量小，发光本领低，所以吸积盘远没有星系的其他部分亮。这样，天文学家通过光学望远镜只看到了星系，没看到吸积盘。不过，吸积盘、旋转黑洞和穿过黑洞的磁力线也许会像图9.7（d）（布兰福德-茨纳耶克过程）那样共同产生强烈的喷流，喷流从星系流出来，进入星系际空间，在那里将能量传给星系的巨大射电叶。

这些以黑洞为基础的关于类星体和射电星系的解释是很成功的，人们不禁会说，它们一定是正确的，而星系的喷流一定就是那个向我们叫喊“我来自黑洞”的惟一信号！不过，天体物理学家还是真够谨慎的，他们更相信铁一般的事实。而所有这些射电星系和类星体的观测性质都可能有另一种不需要黑洞发动机的解释：那可能是一颗比太阳重几百万或几十亿倍的超大质量的磁化的快速旋转的恒星——这种恒星，天文学家从未见过，但从理论上看，它可以在星系中央形成。这样的超大质量恒星的行为很像一个黑洞的吸积盘，它通过收缩（但不能收缩到临界周长以下）可以释放大量的引力能；引力能通过摩擦为恒星加热，使它像吸积盘那样发亮；束缚在恒星的磁力线可以旋转，以喷流形式将等离子体抛出去。

某些射电星系和类星体的能源可能就来自这种超大质量恒星。但是，物理学定律认为，这样的恒星会逐渐缩小，并在接近临界周长时发生坍缩而形成黑洞。恒星在坍缩之前的总寿命会远小于宇宙的年龄。这意味着，虽然最年轻的射电星系和类星体可能以超大质量恒星为能源，但更老的能源，几乎肯定来自巨黑洞——几乎肯定，但不是绝对肯定。这些论证还没成为铁的事实。

巨黑洞多吗？在80年代，逐渐积累的证据表明，这样的黑洞不仅存在于大多数类星体和射电星系，也存在于很多大的正常（不射电的）星系，如银河系和仙女座，甚至还存在于某些小的星系核中，如仙女座的矮伴星M32。在正常星系（银河系、仙女座、M32）中，黑洞周围可能没有吸积盘，或者只有很薄的盘，只能流出很少的能量。

我们银河系里这种黑洞的证据（如1993年的）是很诱人的，但还远没有证实。[30]关键的一点证据来自星系中心附近气体云的轨道运动。伯克利加利福尼亚大学的汤斯（Charles Townes）和他的同事们发现，气体云在绕着一个约300万个太阳质量的天体转动；射电观测表明，在这个中心天体位置上有一个很特别的但不太强的射电源——这个射电源惊人地小，还没有我们的太阳系大。不过，这正是我们希望的一个安静的只有薄吸积盘的300万个太阳质量的黑洞应该具有的观测性质，但它们也容易用别的办法来解释。[3]

巨大黑洞可能存在而且存在于星系的中心，这令天文学家感到非常惊奇。不过，现在想来，我们还是容易理解，这样的黑洞怎么能在星系的中心形成。

在任何星系里，当两颗恒星互相经过时，引力会使彼此偏转，脱离原来路径的方向。（NASA的飞船在遇到木星那样的行星时，也会因为这个作用而改变轨道。）由于这个过程，通常有一颗恒星会偏向星系中心，而另一颗会偏离中心。过程累积的结果是，星系中的某些恒星被驱到星系中心。同样，以后会发现，星系内星际气体的摩擦效应，最终会使大量气体落入星系中心。

随着越来越多的气体和恒星汇聚到中心，它们形成的集团的引力也会越来越强。最后，集团引力将超过它的内部压力，坍缩形成一个巨大的黑洞。另一种可能是，集团内的大质量恒星坍缩形成一些小黑洞，这些小黑洞相互碰撞，也与恒星和气体碰撞，从而形成更大的黑洞，最终形成一个统治中心的巨大黑洞。通过估计坍缩、碰撞和联合

等过程所需要的时间，我们可以合理（尽管还不能令人信服）地认为，大多数星系在很久以前就在它们的中心生成了巨黑洞。

假如不是天文观测令人强烈感到星系的中心存在着巨黑洞，天体物理学家可能在90年代的今天也不会预言它的存在。不过，观测的确令人想到巨黑洞，天体物理学家也很容易让自己适应这种想法。从这一点可以看到，对星系中心真正发生了什么事情，我们的认识是多么贫乏。

未来会怎样呢？我们需要担心银河系里的巨黑洞会吞噬地球吗？看几个数字，我们就可以放心了。我们星系中央的黑洞质量（如果确实存在的话）是太阳的300万倍，于是有5 000万千米或200光秒的周长——大约是地球绕太阳的轨道周长的十分之一，同银河系本身的大小相比，这是很小的。我们的地球跟着太阳一起在一个20万光年周长的轨道上绕着星系中心转动——那比黑洞的周长大300亿倍。假如这个黑洞最终会吞噬银河系的大部分物质，它的周长也只能扩张到1光年左右，我们的轨道周长还比它大20万倍。

当然，在$10^{18}$年里——这是我们的中央黑洞吞噬大部分星系物质所需要的时间（比宇宙现在的年龄还大1亿倍），地球和太阳的轨道也许会发生根本的改变。我们不可能预知这些改变的细节，因为我们不能充分地知道在这$10^{18}$年里太阳和地球可能遇到的其他恒星的位置和运动情况。这样，我们不可能预知太阳和地球最终是会落入星系中央的黑洞，还是会被抛出银河系。然而我们可以相信，即使地球最终会被吞噬，那也是在大约未来的$10^{18}$年——在那遥远的日子来临前，几乎可以肯定会有别的灾难同时降临地球和人类。[4]

[1] 由于星际和星体的气体运动，宇宙在时刻不断地产生磁场；磁场一旦产生，就很难消失。星际气体汇聚到吸积盘时，将自身的磁场也带来了。

[2] “类星体”（quasar）就是“类似于星体的天体”（quasi-stellar）的简称。

[3] A.Eckart和R.Genzel近年通过对银河系中心0.3光年内的星体三维速度的研究，证

明了银河系中心有一个250万个太阳质量的黑洞。——译者

[4] 作者似乎忘了，太阳系的（当然也包括地球的）寿命大约是100亿年，而现在已经过了46亿年，即使没有“天外来祸”，再过50亿年，太阳系自己就可能发生“根本的改变”。——译者

# 第10章　曲率波

引力波把黑洞碰撞的交响曲带给地球，

物理学家设计出仪器来寻找那些波，

倾听它们的音乐

## 交响

10亿年前，在离地球10亿光年远的一个星系的中心，紧密聚集着一个几亿颗恒星和气体的集团。当一颗颗恒星被抛出去后，留下的1亿颗恒星落向中心，集团逐渐收缩，1亿年后，收缩到几光年大小，小恒星也开始零星碰撞、结合，形成更大的恒星。大恒星燃尽它们的燃料，然后坍缩形成黑洞；一对对的黑洞相互靠近，有时落入对方的轨道。

图10.1画了这样一个黑洞的双星系统的嵌入图。每个黑洞在嵌入的表面上形成深坑（强大的时空曲率），当黑洞相互绕着对方旋转时，转动的坑产生曲率波，以光速向外传播。波动在黑洞周围的时空体形成螺旋的波纹，那样子很像花园里高速旋转的浇水器喷出的水。每一滴水都会近似地沿半径飞出，同样，每一点曲率也都会向外辐射开去；向外飞出的水滴形成一条螺旋水线，因此，所有的曲率波也在时空体上形成螺旋的峰谷。



图10.1　两个黑洞组成的“双星系统”轨道“平面”空间曲率的嵌入示意图。中心的两个坑代表黑洞周围的强烈时空弯曲。这样的坑在以前的黑洞嵌入图（如图7.6）中已经遇见过了。当黑洞互相围绕对方转动时，会产生向外传播的曲率波动，叫引力波。[加利福尼亚理工学院LIGO计划提供。]

因为时空曲率与引力是同一件事，所以这些曲率的波动实际上就是引力的波动，或者说引力波。爱因斯坦的广义相对论不容争辩地预言，当两个黑洞或者两颗恒星相互绕着对方转动时，一定会产生这样的引力波。

向外面空间传播的引力波会对黑洞产生反冲，就像射出去的子弹对枪的反冲一样。波的反冲作用使黑洞靠得更近，转得更快；也就是说，黑洞将螺旋式地慢慢落向对方。这个过程会逐渐释放引力能，一半进入引力波，另一半提高黑洞的转动速度。

黑洞的螺旋式运动先很慢，但随着它们越靠越近，会越动越快，它们辐射出的曲率波越强，失去的能量也越多，而螺旋式下落也越快（图10.2（a），（b））。最后，当每个黑洞接近光速时，它们的视界便在接触中结合在一起。原来有两个黑洞的地方，现在只有一个——快速旋转的哑铃型黑洞（图10.2（c））。当视界旋转时，哑铃

辐射出曲率波，波反作用在黑洞上，将哑铃的凸起一点点削去（图10.2（d）），留下一个赤道断面完全光滑而圆的旋转的黑洞视界，正好是爱因斯坦场方程的克尔解所描述的形状（第7章）。

我们不可能有什么办法从最后这个光滑的黑洞发现它的历史，也不可能区分它是两个小黑洞聚合形成的，还是一颗物质恒星或者一颗反物质恒星直接坍缩形成的。黑洞没有能泄露它历史的“毛”（第7章）。

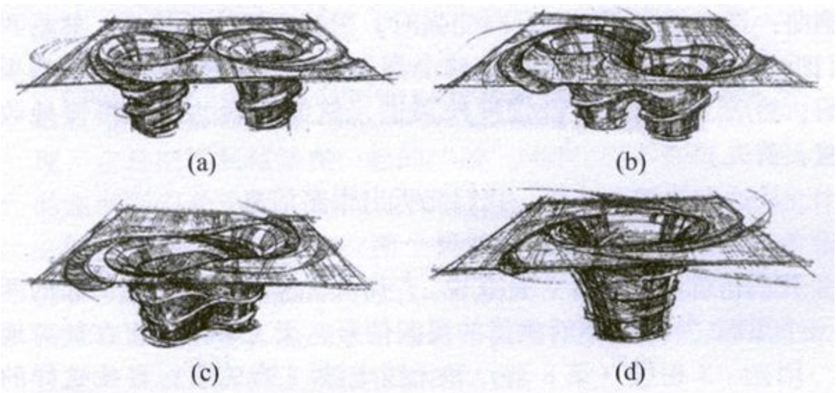



图10.2　两个黑洞组成的“双星系统”周围空间曲率的嵌入示意图。图经过艺术家修饰，看起来很有动感。两个黑洞螺旋式地靠近，这些图表现了这个时间序列。在图（a）和（b），黑洞视界在坑底还是两个圆，它们在图（c）前消失而形成单独的一个哑铃型的视界。旋转的视界发出引力波，也带走了变形，留下一个光滑旋转的克尔黑洞，见图（d）。[加利福尼亚理工学院LIGO计划提供。]

然而，历史没有完全失去，还留下一点儿记录：记录在黑洞结合所发出的时空曲率波里。这些曲率波很像交响音乐的声波。音乐的交响表现在声音的强弱短长（这儿声音大，那儿声音小；这儿频率高，那儿频率低）；黑洞结合的历史也表现在曲率波的大小高低。声波带着交响乐从乐队流向听众；曲率波也带着它的历史从结合的黑洞飞向遥远的宇宙。

曲率波从两个黑洞诞生的恒星和气体的集团里出来，在时空体里

穿行，既不会被吸收，也不会受干扰，完好地保留着历史的记忆。它穿过自己的星系，进入星系际空间；穿过它的星系所在的星系团，然后穿过一个又一个的星系团，来到我们的星系团，我们的银河系，我们的太阳系，最后穿过我们的地球，继续飞向更遥远的星系。

聪明的人类应该能在这些时空曲率波经过时监测到它们；我们的计算机可以将这些曲率波转换成声波，让我们听到黑洞的交响曲：当它的音调逐渐升高变强时，黑洞在螺旋地接近；然后我们听到它疯狂的回旋，那是两个黑洞正在结合成一个变形的黑洞；然后，它拖着长音慢慢地消逝，就像黑洞的凸起慢慢地收缩、消失。

从这支波澜交响曲，我们能听出很多信息：

1.我们仿佛听到一个声音在说，“我来自螺旋式结合在一起的两个黑洞。”这是绝对确凿的黑洞信号，天文学家一直在徒劳地用光、X射线（第8章）和无线电波（第9章）寻找这样的信号。因为光、X射线和无线电波在远离黑洞视界的外面，从一类完全不同于黑洞组成（纯时空曲率）的物质（高速热电子）产生出来，在穿过中间物质时会遭受严重的破坏，所以它们不能携带多少关于黑洞的信息，更不可能有什么确定的信号。相反，曲率波（引力波）来自结合黑洞视界的邻近，是由与黑洞同样的物质（时空结构的弯曲）产生的，不会遭受传播途中物质的破坏，所以，它们能为我们带来具体的关于黑洞的消息和确凿的黑洞信号。

2.这支波澜交响曲能告诉我们，每个黑洞有多重，它们旋转有多快，它们的轨道是圆还是直，它们在天空什么地方，它们离地球有多远。

3.交响曲还表现了螺旋黑洞的时空曲率的部分特征，我们能第一次确定性地检验广义相对论的黑洞预言：交响曲所表现的图景与爱因斯坦场方程的克尔解一致吗（第7章）？它所表现的旋转黑洞附近的

旋涡是克尔解要求的吗？旋涡的数量与克尔解的相同吗？旋涡在接近视界时像克尔解说的那样变化吗？

4.交响曲还表现了两个黑洞视界的结合和结合的黑洞的振荡——这些事情我们今天也只有很模糊的认识。因为爱因斯坦广义相对论定律与它们相关的那个特征，我们还理解得太少，那就是所谓非线性的特征（卡片10.1）。“非线性”意味着大曲率本身还要产生曲率，它反过来又产生更大的曲率——像雪崩，下滑的一点儿雪带动周围的雪，它们又带动更多的雪，最后一坡的雪都滑落下来。我们认识非线性在宁静黑洞的表现，它是把黑洞粘结在一起的“胶”。但我们不知道，当强大的曲率剧烈动荡时，非线性在做什么：它如何表现？产生什么效应？为了认识它们，两个黑洞的结合与振荡是很有希望的“实验室”。为了认识它们，还需要实验物理学家和理论物理学家并肩协作，监测来自遥远宇宙的结合黑洞的交响波澜，在超大规模计算机上模拟它们的结合。

卡片10.1　非线性及其结果

如果一个量的总体是部分之和，我们就称它是线性的；否则，它就是非线性的。

我的家庭收入是线性的：它是妻子和我自己的薪水之和。我退休以后的养老金是非线性的，它不是我过去投入的总和；相反，它远比那个和大，因为每一笔投入都有利息，而每一点利息又会为自己带来利息。

下水道的水量是线性的，它是每家倒进管道的水的总和。雪崩的体积是非线性的，一点儿雪能够诱发一山坡的雪崩落。

线性现象简单，好分析，好预测。非线性现象复杂，难分析，难预测。线性现象只有很少的几种行为方式，很容易分门别类。非线性现象五花八门——科学家和工程师们近几年在遇到被称为混沌的非线

性行为时，才开始认识它们。（混沌思想的优美引导，请看格莱克（Gleick）1987年的书。[1]）

时空曲率小（如在太阳系）时，近似为线性的，例如，地球上海洋潮汐就是月亮和太阳的时空曲率（潮汐引力）联合作用的结果。相反，时空曲率大（如在大爆炸或黑洞附近）时，爱因斯坦广义相对论引力定律预言，曲率是高度非线性的——是宇宙中极端非线性现象之一。然而，目前我们几乎还不能说明引力非线性特征的实验和观测数据，我们解爱因斯坦方程的能力还低得可怜，我们的解只有在很简单的情况下——例如，在宁静的旋转黑洞附近，才说明了一点非线性的东西。

宁静黑洞因为引力的非线性而存在；离开引力的非线性，黑洞自身都不能维持，就像木星上的大红斑，如果没有气体的非线性行为，也不能存在下去。当生成黑洞的坍缩恒星消失在黑洞视界里时，它也失去了以任何方式影响黑洞的能力；最重要的是，恒星的引力不再是黑洞的维持者。这时候，黑洞还能继续存在完全是因为引力的非线性：没有了恒星，黑洞时空曲率仍将继续产生其非线性；这样自我生成的曲率像非线性“肢”一样将黑洞粘在一起。

宁静黑洞激起了我们的兴趣，我们也还想知道更多：引力的非线性还产生了别的什么现象吗？通过监测和解读结合黑洞所产生的时空曲率波，也许能得到一些答案，我们在那儿大概会遇到从没想到过的混沌和奇异行为。

为了认识它们还需要监听曲率的交响乐。怎么听呢？关键在于曲率的物理本质：时空曲率与潮汐引力是同一回事。月亮产生的时空曲率在地球上激起海潮（图10.3（a））；同样，引力波的时空曲率也能激起海潮（图10.3（b））。

然而，广义相对论认为，月亮激起的海潮与引力波激起的海潮有三点不同。第一点，传播不同。引力波的潮汐力（曲率波）类似于光

波和无线电波，它们从源地以光速传向地球，在传播中振荡。而月亮的潮汐力像带电物体的电场。电场紧紧依附在带电体上，带着电场的物体就像刚毛耸立的刺猬；月亮的潮汐力也是这样的，像从月亮伸出数不清的手，随时准备捕获、挤压或拉伸走近它的事物。月亮的潮汐力在地球海洋中引起的潮涨潮落似乎每过几个小时就会改变，那只是因为地球在引力场中转动。假如地球不转动，潮汐力的作用是不会改变的。

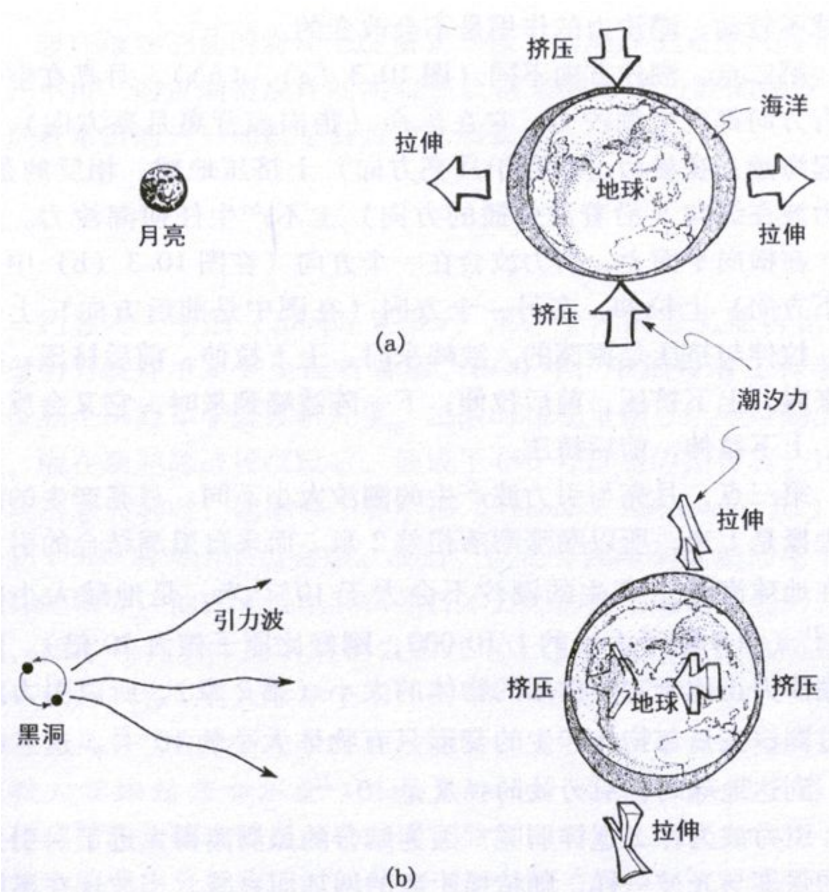



图10.3　月亮和引力波产生的潮汐力。（a）月亮的潮汐力在地球上引起的海洋潮汐的涨落：纵向涨潮，横向落潮。（b）引力波的潮汐力引起海洋潮汐的涨落。力完全是横向的，在一个方向上拉伸，另一个方向上挤压。

第二点，潮汐方向不同（图10.3（a），（b））。月亮在空间所有方向都产生潮汐力。它在纵向（指向或背离月亮方向）上掀起海潮，在横向（垂直于月亮方向）上挤压地球。相反的是，引力波在纵向（沿着波传播的方向）上不产生任何潮汐力。然而，在横向平面上，引力波会在一个方向（在图10.3（b）中是上下方向）上拉伸，在另一个方向（在图中是前后方向）上挤压。拉伸与挤压是振荡的，波峰来时，上下拉伸，前后挤压；波谷来时，上下挤压，前后拉伸；下一阵波峰到来时，它又会反过来，上下拉伸，前后挤压。

第三点，月亮与引力波产生的潮汐大小不同。月亮产生的潮汐大概是1米，所以潮涨潮落相差2米。而来自黑洞结合的引力波在地球海洋上产生的潮汐不会大于$10^{-14}$米，是地球大小的$10^{-21}$（单个原子大小的1/10 000，刚好比原子核大10倍）。因为潮汐力正比于它所作用的物体的大小（第2章），所以引力波通过潮汐在任何物体产生的变形只有物体大小的$10^{-21}$。这意味着，到达地球时，引力波的强度是$10^{-21}$。

引力波为什么这样弱呢？因为结合的黑洞离得太远了。引力波的强度与光波一样，随传播距离的增加而衰减。当波还在黑洞附近时，它们的强度大概为1，就是说，物体有多大，它就把它拉伸或压缩多少，在这么强的作用下，人是会丧命的。然而，到达地球时，引力波的强度减小到约（1/30黑洞周长）/（波经过的距离）。[2]对一个10亿光年远、10个太阳那么重的黑洞来说，引力波强度为（1/30）×（180公里的黑洞周长）/（到地球的10亿光年）≈$10^{-21}$。于是，它使地球海洋发生的形变为$10^{-21}$×（$10^{7}$米的地球大小）=$10^{-14}$米，正好是原子核直径的10倍。

想在地球汹涌的海洋上测量这么微小的潮汐是完全没有希望的。不过，通过周密设计的实验室仪器来测量引力波的潮汐力，还是有希望的——那就是引力波探测器。

[1] James Gleick，Chaos，Making a New Science。这本书至少已经有3个中译本了，例如张淑誉译、郝柏林校，《混沌，开创新科学》（上海译文出版社，1990）。——译者

[2] 因子1/30来自爱因斯坦场方程的具体计算。它包括一个将黑洞周长换算成半径的因子1/2π，近似等于1/6；另一个因子1/5是爱因斯坦场方程决定的。

# 棒

约瑟夫·韦伯（Joseph Weber）第一个充满远见地认识到，测量引力波并不是完全没有希望。1940年，韦伯带着工程学士学位到美国海军学院读研究生，二战时在勒星顿号航空母舰上服役，舰在珊瑚海战役沉没后，他成了690号潜艇的指挥官。1943年攻占意大利时，他率领小罗斯福（Theodore Roosevelt，Jr.）准将和1 900名突击队员登岸。战后，他是美国海军舰船局电子对抗部的领导。他在无线电和雷达技术方面的造诣是出了名的，所以，1948年马里兰大学请他去担任电子工程学教授，他同意了——成为一名只有大学学士学位的29岁的正教授。

在马里兰大学讲电子工程的同时，韦伯也在准备改行：在天主教大学跟赫兹菲尔德（Karl Herrfeld）学成了物理学博士。这位赫兹菲尔德也曾是惠勒的博士导师。韦伯从他那儿学到了很多关于原子、分子和辐射的物理学，从而在1951年发现激光产生的新机制，但他没有条件用实验来证明他的思想。在韦伯准备发表他的思想时，[1]另有两个小组——一个在哥伦比亚大学，由汤斯领导；一个在莫斯科，由巴索夫（Nikolai Gennadievich Ba-sov）和普罗哈洛夫（Aleksandr Michailovich Prokharov）领导——独立发现了另一种新机制，而且他们接着就实现了激光。[1]

尽管韦伯关于这个机制的论文最先发表，但他几乎什么荣誉也没得到；诺贝尔奖和专利都给了哥伦比亚和莫斯科的科学家。[2]韦伯很失望，但他与汤斯和巴索夫仍然很友好。他又去找新的研究方向。

韦伯先找到了惠勒的研究小组，参加了一年，成为一名广义相对论专家，跟惠勒一起研究广义相对论对引力波性质的预言。1957年，他又找到了新方向，开始建造世界上第一台寻找和监测引力波的仪

器。

从1957年下半年开始，到1959年初，韦伯想尽了他能想到的各种引力波探测方案。这还只是在纸上的脑力训练，没做实验。他的想法写满了4本300页的笔记，包括可能的探测器设计和每一设计预想步骤的计算。一个个想法都因为粱有希望而被他扔到一边。但有几个还有希望，韦伯最后选择了一根圆柱形铝棒，长约2米，直径0.5米，重1吨，侧面指向到来的引力波（图10.4下）。[2]

引力波因为潮汐力的振荡，将交错挤压和拉伸铝棒的两端。铝棒两端具有相对于中心向里和向外振动的自然振动模式，能与振荡的潮汐力发生共振。这种自然振动模式就像铃铛、音叉或者酒瓶的声音一样，有非常确定的频率，与这些自然频率相应的声波能使这些物体产生共鸣；同样，与棒的自然频率相应的振荡潮汐力也能引起棒的共振。于是，以这种棒作为引力波探测器，应该将棒的大小调整到使它具有与到来的引力波相应的自然频率。

频率该是多少呢？1959年韦伯开始他的计划时，相信黑洞的人很少（第6章），相信者也只认识很少一点儿黑洞性质。那时还没人想到黑洞会碰撞、结合并发射记录它们碰撞历史的时空曲率波，也没人能就其他的引力波源提出有多大希望的指导。

所以，韦伯是从黑暗中摸索起步的。他惟一的指南是粗略（却是正确的）知道引力波的频率大概低于10 000赫兹（每秒转10 000周）——那是物体以光速绕最紧致的恒星，即接近临界周长的恒星运动的轨道频率。[3]于是，韦伯设计了他能做到的最好探测器，让它们的共振频率尽可能都落在10 000赫兹以下，希望宇宙也能提供具有他所选择的频率的波。很幸运，他的铝棒的共振频率大约是1 000赫兹（每秒振荡1 000周），后来发现来自结合黑洞的某些波正好就在这样的频率上振荡。某些来自超新星爆炸和结成的中子星对的引力波，也是这种频率。



图10.4　韦伯在介绍绕着铝棒中央贴上去的压电性晶片（约1973年）。引力波驱动铝棒两端振动，振动将内外挤压这些晶片，从而产生可以用电学方法检测的电压。[James P.Blair摄，国家地理学会提供。]

韦伯计划里最困难的地方在于发明一个用来监测铝棒振动的传感器。他料想，波产生的振动应该很小，小于一个原子核的直径（但在60年代，他不知道那到底是多小。据最近的估计，它只有$10^{-21}\times$（2米棒长）$\approx 10^{-21}$米，或者说，只有原子核直径的百万分之一）。对五六十年代的大多数物理学家来说，即使原子核直径的十分之一也是不太可能测量的。但韦伯不这样看，他发明了能胜任的传感器。

韦伯传感器的基础是压电效应：某些类型的材料（特定的晶体或陶瓷）在受轻微挤压时会在两端产生电压。韦伯本想用这类材料来做

他的棒，但材料太贵，他只好求其次：用铝做棒，然后绕着棒的中央贴上一些压电性晶体片（图10.4）。棒振动时，表面将挤压或拉伸晶片，每块晶片都产生一个振荡电压。韦伯用电路将晶片一块块串联起来，于是微弱的振荡电压将叠加在一起，即使棒的振动只有原子核直径的十分之一，这样累积的电压也足以用电学方法检测出来。

60年代初，韦伯还是世界上惟一一个寻找引力波的实验物理学家。带着激光竞争的痛苦回味，他喜欢这种孤独。然而，在70年代初，他那令人感动的敏锐力和他实际可能检测到了的引力波证据（现在想来，我相信他没有检测到），吸引了几十位实验家；80年代，已经有100多位有才能的实验家投身进来，为实现引力波的天文学与他并肩战斗。[4]

我第一次见韦伯是在法国阿尔卑斯山勃朗峰对面的山坡上，那是1963年的夏天，他着手探索引力波已经4年了。那时我还是刚开始研究相对论的研究生，和来自世界各地的其他35名学生一起到阿尔卑斯山来参加紧张的两个月的爱因斯坦广义相对论引力定律的暑期讲习班。[5]老师都是世界上最伟大的专家——惠勒、彭罗斯、米斯纳、德维特（Bryce Dewitt）、韦伯等等——我们在课堂上听他们讲，私下里与他们交谈。南山的雪闪着耀眼的光芒，勃朗峰直插我们头上的天空；在我们周围，牛群带着铃响在绿油油的牧场上吃草，山下离学校几百米的地方，是美丽如画的莱苏什的村庄。

在这迷人的地方，韦伯讲引力波和他的探测计划，也令我人迷了。课后，我与韦伯谈物理、谈生活，也谈登山。逐渐发现，他和我个性相投。我们都喜欢独处，不喜欢紧张的竞争和激烈的思想讨论。我们更喜欢自己考虑问题，偶尔从朋友那儿听一些建议和想法，但又不让想赶在我们前头获得认识和发现的人从我们身边超过去。

接下来的10年里，黑洞研究热起来了，进入了它的黄金年代（第7章）。我开始感到黑洞研究并不令人喜欢——它太紧张、太激烈、

太混乱。于是我忙着找别的自由空间多一些的研究领域，那样我在投入大部分精力后，还可以有点儿时间来研究黑洞或别的事情。在韦伯激发下，我选择了引力波。

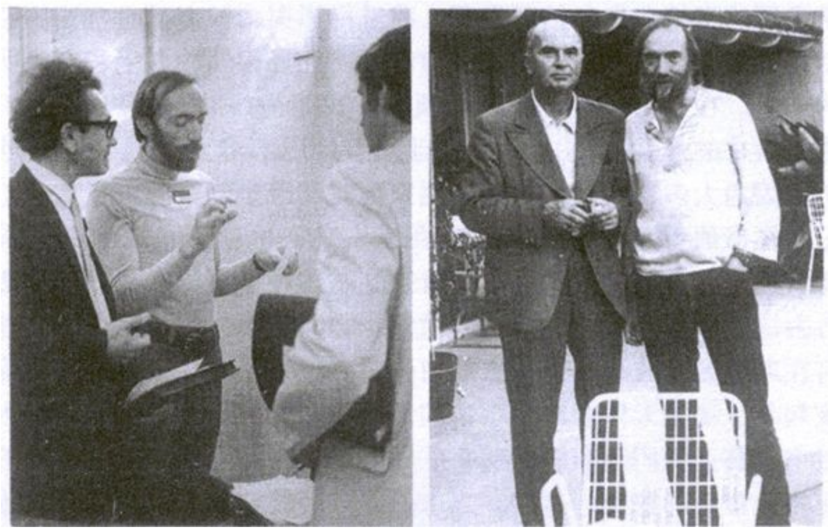
我同韦伯的看法一样，引力波的研究还很幼稚，但它有光明的未来。在它萌芽时走进这个领域，我能为帮助它成长而快乐，能为后来的建设者奠定一点基础，而且用不着别人在我耳边唠叨，因为大多数其他相对论理论家那时都聚集到黑洞去了。

在韦伯看来，需要的基础在于实验，也就是，探测器的设计、建造和不断的改进。在我看来，基础在于理论。我们应该努力去认识，爱因斯坦的广义相对论定律关于引力波如何产生、如何在离开时对波源反作用，如何传播，都说了些什么；我们还应该判断，哪类天体会产生宇宙间最强的引力波，有多强，以多大频率振荡；我们还应该发明一些数学工具来解开这些天体产生的交响曲背后的秘密，这样，当韦伯等人最终探测到引力波时，理论和实验才能进行对比。

1969年，应泽尔多维奇的邀请，我在莫斯科过了6个星期。泽尔多维奇向我和其他一些人讲了他的一大堆新想法（第7章、12章）。一天，他抽时间开车送我去莫斯科大学，把我介绍给一位年轻的实验物理学家布拉金斯基（Vladimir Braginsky），他在韦伯激发下发展引力波探测技术已经好几年了，是继韦伯之后最先进入这个领域的实验家。他也做其他有趣的实验，例如，寻找夸克（质子和中子的基本构成物质），检验爱因斯坦关于所有物体（不论它的组成如何）在引力场中以相同加速度下落的论断（这是爱因斯坦将引力描述为时空曲率的基础）。

布拉金斯基给我留下了很深的印象。他机敏而深刻，对物理学有非常好的感觉；他热情而直率，很容易同他谈政治和科学。我们很快成了好朋友，也学会了尊重彼此的世界观。对我这样的美国自由民主党人来说，个人的自由是高于一切的，政府没有权力叫人怎么生活。

对布拉金斯基这样的非教条共产主义者来说，个人对社会的责任才是高于一切的。



左：1973年9月，韦伯、索恩和Tony Tyson在波兰华沙的一次引力辐射会议上。右：1984年10月，布拉金斯基和索恩在加利福尼亚帕萨迪纳。[左，Marek Holzman,Andrzej Trautman提供；右，Valentin N.Rudenko提供。]

布拉金斯基具有别人没有的远见。在我们1969年见面时，以及后来在1971年和1972年再见时，他都警告我，寻找引力波的棒存在着一个根本的最终极限。[6]他告诉我，那个极限来自量子力学。尽管我们一般认为量子力学只对电子、原子和分子那样的小事物发生作用，但是，如果对1吨棒的振动测量足够精确，我们会发现那些振动也有量子力学行为，而且这些行为最终会给引力波的探测带来问题。布拉金斯基很相信这一点，因为他计算过韦伯的压电性晶体和其他几类可能用于棒的振动测量的传感器的最终行为。

我不明白布拉金斯基在说什么。我不懂他的理由，不懂他的计算，也不懂它的重要性，所以没太注意。他向我讲的其他事情似乎要重要得多：我从他那儿学会了如何考虑实验，如何设计实验装置，如何预测影响仪器的噪声，如何消除噪声使仪器正常运行——而布拉金

斯基从我这儿学的是，如何认识爱因斯坦的引力定律，如何确认那些预言。我们很快结成一个小组，每个人都把自己的专业带进我们共同的事业。在接下来的20年里，我们将得到巨大的快乐，也有一些发现。

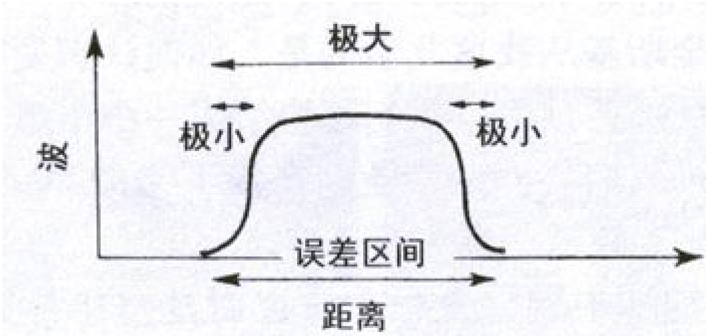
70年代初期和中期，我和布拉金斯基每年都见面，在莫斯科、帕萨迪纳、哥本哈根、罗马或别的什么地方，他每年都警告我量子力学会给引力波探测器带来麻烦，而我每次都没听明白。他的警告有些乱，因为他自己也没完全明白发生了什么事情。然而到1976年，斯坦福大学的吉法德（Robin Giffard）在布拉金斯基后也独立提出这样的警告，而且说得更清楚，我才恍然大悟。我终于意识到问题的严重：棒探测器的最终灵敏度严格受测不准原理的限制。[7]

测不准原理是量子力学的一个基本特性。它说的是，如果你想高度精确地测量一个物体的位置，那么在测量过程中，你必然会对物体有一种反作用，从而以一种随机的不可预料的方式干扰物体的速度。位置测量越精确，物体速度受到的不可预料干扰就越强烈。不论仪器设计得多么巧妙，你都不可能超越这种固有的不确定性（见卡片10.2）。

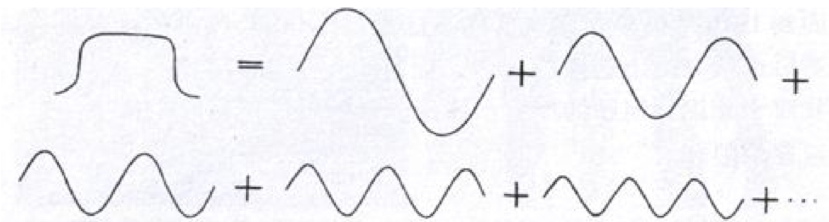
卡片10.2　测不准原理与波粒二象性

测不准原理与波粒二象性（卡片4.1）——也就是粒子有时像波、有时像粒子的行为趋向——是密切相关的。

假如你在测量一个粒子（或者别的物体，如棒的端点）的位置，确定它在某个误差区间，那么，不论粒子的波在测量前像什么，在测量中，测量仪器都会对它产生反作用，从而将它约束在误差区间内。于是，得到的波形有点儿像下面的样子：




这样的约束波包含了从误差区间本身大小（图上标极大）到波的两端所在小区域的大小（图上标极小）的不同波长。更具体地说，受约束的波可以通过下面这些波长从极大到极小的波动之和或叠加来构成：



现在，想象波振荡的波长更短，粒子的能量更大，从而粒子的速度也更大。因为测量为波限定了一个波长范围，所以粒子的能量和速度也一定落在一个相应范围内；换句话说，它的能量和速度是不确定的。

概括地讲，测量将粒子的波约束在一定误差区间内（上面第一个图），使波由一定范围内的波长组成（第二个图），而波长的范围对应一个能量和速度的范围，从而速度是不确定的。不论你费多大气力，你在测量粒子的位置时，都免不了产生这种速度的不确定性。而且，更深入的论证表明，位置测量越精确，即误差区间越小，波长和速度的范围就越大，这样，粒子速度的不确定性也就越大。

测不准原理不仅决定电子、原子和分子等微观事物的测量，也影响宏观事物的测量。但是，由于大物体有大惯性，测量的反作用只能产生很轻微的速度扰动。（速度受到的干扰与物体质量成反比。）

在引力波探测器问题上，测不准原理说的是，传感器对振动棒两端的位置测量越精确，测量对棒产生的随机反作用就越强大。

对于不精确的传感器，测量的反作用可能很小而无关紧要，但如果传感器不精确，你从哪儿知道棒的振动幅度呢？当然也就更不可能监测到微弱的引力波。

对于极端精确的传感器，反作用可能很大，能强烈改变棒的振动。这些巨大的未知的变化，将淹没你想探测的任何引力波。

在这两个极端之间，存在一个理想的传感器精度：它既不因为太低而令你一无所获，也不因为太高而出现不可知的强大反作用。在这样一个现在称为布拉金斯基标准量子极限的理想精度下，测量产生的反作用的效应几乎与传感器产生的误差一样小。没有传感器能比这个标准量子极限更精确地监测棒的振动。那么极限是多大呢？对2米长、1吨重的棒来说，大约比一个原子核小100 000倍。

60年代，谁也没有认真考虑过需要这么精确的测量，因为没人清楚地知道来自黑洞和其他天体的引力波会有多么微弱。不过到70年代初，在韦伯实验计划激励下，我和其他理论家已经指出了最强引力波可能具有的强度，大约是$10^{-21}$，[8]这意味着波在2米棒产生的振动幅度只有$10^{-21}$×（2米），约一个原子核直径的百万分之一。如果这些估计是正确的（我们也知道那是很不确定的），那么引力波信号比布拉金斯基标准量子极限小10倍，从而不可能用棒和任何已知类型的传感器来监测。

这实在令人忧虑，但并不是一切都完了。布拉金斯基深刻的直觉

告诉他，如果实验者有特别机灵的办法，还是可能超越他的标准极限。他指出，这需要用一种新办法来设计传感器，使它不可避免的未知的反作用不会掩盖引力波对棒的影响。布拉金斯基称这样的传感器为量子无破坏[3]传感器。“量子”是由于传感器的反作用来自量子力学定律的要求，“无破坏”说的是传感器的设计避免了反作用对被测物体的破坏，也就是反作用不会破坏引力波对棒的影响。布拉金斯基也没有可行的量子无破坏传感器的设计，但直觉告诉他，这样的传感器应该是可能的。

这一次我很认真地听了布拉金斯基的话；在接下来的两年里，我和我在加州理工学院的小组以及他和他在莫斯科的小组都在断断续续地努力，为的就是设计一台量子无破坏传感器。

1977年秋，我们同时找到了答案——但方法完全不同。[9]我清晰记得，我当时是多么兴奋。那是某一天在格里西（学院的学生食堂）午餐后，凯维斯（Carlton Caves）和我在激烈讨论中突然想到的。[4]我还记得，当我得知布拉金斯基、沃罗索夫（Yuri Vorontsov）和哈里利（Farhid Khalili）几乎同时在莫斯科发现了相同的重要思想时，我心中涌起一股辛酸和喜悦的感觉——辛酸是因为我曾满以为自己是新事物的第一个发现者；喜悦是我为布拉金斯基感到骄傲，为能和他同享一个发现而感到高兴。

我们的量子无破坏思想很抽象，它允许很大一类传感器设计超过布拉金斯基的标准量子极限。然而，因为思想抽象，我很难解释。所以，我在这儿只讲一个（不太实用的）量子无破坏传感器的例子。[5]布拉金斯基称它是频闪传感器。

频闪传感器依赖于棒振动的一个特殊性质：假如棒受到一个尖锐的未知反冲作用，它的振幅将发生改变，但不论振幅怎么变化，经过一个振动周期后，棒的振动端将回到它受反冲时的位置（图10.5中的黑点）。至少在引力波（或其他力）没有同时作用在棒上时是这样

的。假如引力波（或其他力）同时在挤压棒，那么一个周期后，棒的位置会发生改变。

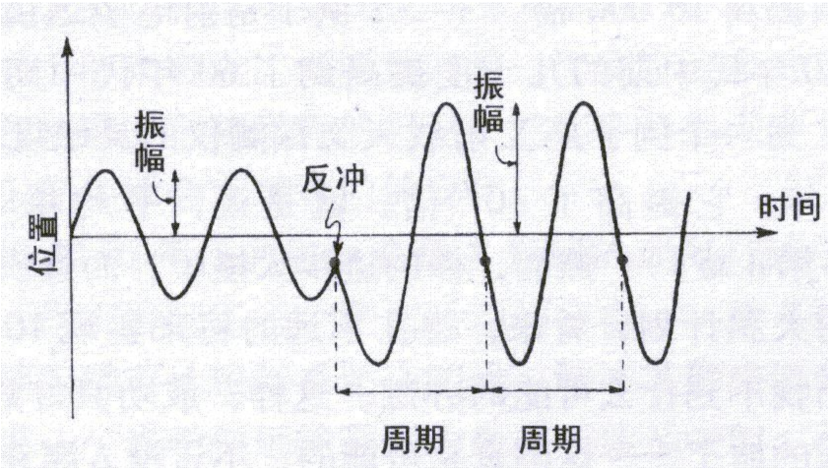



图10.5　频闪量子无破坏测量方法的原理。纵向画的是振动棒端点的位置；横向为时间。如果在反冲时刻迅速而高精度地测量一个位置，则传感器将对棒突然产生一个未知的反冲作用，从而以未知方式改变棒的振幅。然而，在一个两个或者若干个周期后，棒端的位置仍然不会改变，还是与反冲时刻的位置一样，而且完全与反冲作用无关。

于是，为了探测引力波，应该造一台对棒的振动端进行频闪测量的传感器，也就是，传感器在每个振动周期内很快地测一次棒端的位置。这样的传感器在每次测量时都会对棒产生反冲作用，但这些反冲作用在后续测量时不会改变棒端的位置。如果发现位置变了，那么一定有引力波（或其他力）作用在棒上。

虽然量子无破坏传感器克服了布拉金斯基的标准量子极限，但到80年代中期，我却对棒探测器的前景感到悲观，恐怕它不会为引力波天文学带来什么结果。我悲观的原因有两个：

第一，尽管韦伯、布拉金斯基以及其他一些人做的探测棒已经达到了50年代不可想象的灵敏度，但它们只能可信地用来探测强度在$10^{-17}$以上的引力波，如果我和其他一些人对到达地球的引力波强度没

有估计错的话，这个精度离成功还差10 000倍。这本身倒并不严重，因为在20年的时间里，技术的进步常能使仪器的本领提高10 000倍。（一个例子是射电望远镜的角分辨率，它从40年代中期的几十度提高到了60年代中期的几弧秒（第9章）。另一个例子是X射线天文探测仪的灵敏度，从1958年到1978年，它提高了$10^{10}$倍，就是说，平均每8年提高10 000倍（第8章））。然而，棒的进步太慢了，而且没有未来技术和工艺的大胆计划，看来，想在不远的将来实现10 000倍的进步，恐怕找不到什么可能的办法。这样，成功只好靠比$10^{-21}$的估计更强的波了——这倒真是可能的，不过没人愿意依赖它。

第二，即使棒探测到了引力波，要解释它的交响信号也将遇到巨大的困难，实际上很可能会失败。原因很简单：正如音叉或酒杯只对接近其自然频率的声波产生共振，棒也只对接近它自然频率的引力波才有响应；从技术上说，棒探测器只有一个很窄的带宽（带宽就是它产生响应的频率范围），但引力波的交响信号通常混合着一个范围很宽的频率。于是，为了析取这些波的信息，需要一个由许多棒组成的“木琴”，每一根棒覆盖一个不同的小频率带。这架木琴需要多少棒呢？用那时正在计划和制造的那种棒，需要几千根——这实际上是不可能的。原则上讲，要增大棒的带宽，[10]用十几根就够了，但那么做所要求的主要技术进步比达到$10^{-21}$的灵敏度还高。

尽管在80年代，我没有公开讲过多少悲观的话，但我自己还是认为那是可悲的，因为我看到了韦伯、布拉金斯基和我其他朋友和同事为探测棒付出的巨大努力，也因为我已经相信，引力辐射有力量在我们的宇宙认识中产生革命。

[1] 他们的激光器实际产生的是微波（短波长无线电波）而不是可见光，所以被称为脉射（masere），而不是“雷射（lasere）”。“真正的”雷射，也就是产生激光的那种，要几年以后才实现。（maser是microwave amplification by stimulated emission of radiation的缩写，即受激辐射微波放大，或叫微波激射；将微波（microware）换成光（light），就成Laser，即雷射或激光。——译者）

[2]　1964年的诺贝尔物理学奖就是那三位共享的。——译者

[3]　布拉金斯基对英语里的微妙差别有不同寻常的敏感，他比美国人和英国人还更容易造出很有意味的新词来描述他的新思想。（他用的“破坏”是“mondemolition”，在汉语中就难得找到那么微妙的词了。——译者）

[4]　我们思想的重要基础来自英国哥伦比亚大学的同行昂鲁什（William Unruh）。这一思想的发展和结果，主要是因为凯维斯、我和在发现它时与我们同桌进餐的另外三个人：Ronald Drever,Vemon Sandberg和Mark Zimmermann。

[5]　凯维斯等人（1980）以及布拉金斯基、沃罗索夫和索恩（1980）的文章介绍了完整的思想。

# LIGO

为了理解引力波的探测和破译可能带来的革命，让我们先仔细回忆以前的一次革命：由X射线和射电望远镜的进步产生的革命（第8，9章）。

在射电天文学和X射线天文学来临前的30年代，我们的宇宙知识几乎全部来自光。光看到的是一个安宁沉寂的宇宙，充满了恒星和在轨道上平稳运行的行星。它们平稳地发着光，过数百万或者数十亿年才会发现它们的变化。

50、60和70年代的射电波和X射线的观测打破了这种平静的宇宙观，我们看到了一个剧烈活动的宇宙：从星系核喷射出气流，类星体闪耀着比银河系还亮的光，脉冲星射出以光速旋转的强烈辐射束……光学望远镜看到的最亮天体是太阳、行星和少数邻近的宁静恒星。射电望远镜看到的最亮天体是遥远星系中心的猛烈爆炸（能量可能来自巨黑洞）。X射线望远镜看到的最亮天体是从伴星吸积热气体的小黑洞和中子星。

射电波和X射线是因为什么而产生那么壮观的革命呢？关键是，它们给我们带来了比光更多的不同类型的信息。光的波长只有半微米，主要是留在恒星和行星大气中的热原子发出的，所以它为我们带来了关于这些星体的大气的信息。无线电波的波长长1　000万倍，主要是在磁场中近光速螺旋运动的电子发出的，于是，它向我们坦白了星系核射出的磁化喷流，吞没喷流的巨大的星系间的磁化射电叶，以及脉冲星的磁化辐射束。X射线的波长比光短1 000倍，大多数是被吸积到黑洞和中子星的超高热气体中的高速电子发出的，因此它直接反映了黑洞和中子星吸积气体的情况。

一方面是光，另一方面是射电波和X射线，不过，它们之间的差

别同现代天文学的电磁波（光、无线电波、红外线、紫外线、X射线和γ射线）和引力波之间的差别比起来，就是小巫见大巫了。相应地，与射电波和X射线相比，引力波将为我们的宇宙认识带来更大的革命。电磁波与引力波之间的差别，主要是以下这些：[1]

·产生引力波最强烈的应该是时空曲率的大尺度相干振荡（例如，两个黑洞的碰撞和结合）以及大量物质大规模的相干运动（例如，触发超新星的恒星核的坍缩或相互围绕的两颗中子星的螺旋式碰撞和结合）。因此引力波应该向我们展现大曲率大质量的运动。相反，宇宙电磁波通常是单个分离的原子和电子分别发出的，这些以稍微不同的方式振荡的电磁波彼此叠加到一起而形成天文学家观测的波，结果，我们从电磁波得到的主要是发射原子和电子所经历的温度、密度和磁场。

·产生引力波最强的空间区域引力也很强大，在那里，牛顿的描述失败了，应该以爱因斯坦的描述来代替；在那里，大量的物质或时空曲率都近光速地运动、振动或盘旋。例如，宇宙的大爆炸起源、黑洞的碰撞、超新星爆发中心新生中子星的脉冲。由于强引力区域周围通常是厚厚的能吸引电磁波（但不能吸收引力波）的物质层，这些区域不能向我们发射电磁波。相反，天文学家看到的电磁波几乎完全来自弱引力、低速度的区域，如恒星和超新星的表面。

这些差别告诉我们，我们可能用引力波探测器来研究的物体，是不可能通过可见光、无线电波和X射线发现的；而天文学家现在用光、无线电波和X射线研究的物体，它们的引力波也是很难看到的。这样，引力的宇宙和电磁的宇宙看起来会截然不同。我们要从引力波得到的东西不可能从电磁波得到。这也就是为什么引力波可能变革我们对宇宙的认识。

有人会说，现在我们在电磁波基础上对宇宙的认识，比30年代的光学认识要完整得多，将来的引力波革命可能还远不如无线电波和X

射线革命这样壮观。我看不是这样。想起对来到地球的引力波的可怜的估计现状，我就痛苦地感到，我们的认识还差得太远。除了双星和它们的结合以外，对我们考虑过的每一类型的引力波源，不论是这些源的波在离地球一定距离的强度，还是那类源发生的频率（这样也就包括我们到最近源的距离），都存在10的若干次方的不确定性，甚至这些源是否存在也都说不定。

引力波探测器的规划和设计常因这些不确定性而失败，这是令人泄气的；但另一方面，当最终发现并认识了引力波的时候，它可能给我们带来巨大的惊奇。

1976年，我还没有对棒探测器感到悲观，反倒是非常乐观。那时，第一代棒探测器刚有结果，灵敏度比人们预料的好得多；布拉金斯基等人为将来的巨大改进提出了许多灵活而有希望的思想；而我和一些人才刚认识到引力波可能会变革我们对宇宙的认识。

11月的一天晚上，我漫步在帕萨迪纳街头，夜已经很深了，但我心里充满了热情和希望。我在想，是不是该建议加州理工学院设一个引力波探测计划，它的好处是显然的：从一般科学说，如果计划成功了，会带来巨大的精神财富；从学院说，这是占领一个激动人心的新领域的好机会；从我个人说，我可能在自己的学校拥有一个实验家小组，我可以同他们交流，而不再靠地球另一端的布拉金斯基和他的小组了；另外，我可能比往来莫斯科发挥更重要的作用（从而也有更多的乐趣）。但不利因素也是显然的：计划很冒险；为了计划成功，需要学院和美国国家科学基金的大量投入，需要我和其他人付出很多时间和精力；而且，所有这些付出仍然可能失败。这比学院在23年前进入射电天文学的风险要大得多（第9章）。

我独自想了好几个小时，还是没挡住成功的诱惑。经过几个月的风险评估和成果分析，加州理工学院物理学和天文学系及行政部门一致通过了我的建议——但有两个条件。我们得找一位杰出的实验物理

学家来领导这一项目，而且项目要大、要强，以提高成功的机会。就是说，同韦伯在马里兰大学和布拉金斯基在莫斯科以及其他那时正在进行的引力波工作相比，我们需要付出更多更大的努力。

第一步是找领导者。我飞到莫斯科去征求布拉金斯基的意见，也想看看他是不是愿意来。我的话令他心乱了。他面临一系列痛苦的抉择：在美国有好得多的技术，在莫斯科有了不起的工艺（例如，玻璃吹制技术在美购几乎失传了，但莫斯科还有）；在美国，他需要从草稿实现一个计划，在莫斯科，低效官僚的前苏联体制总是疯狂地挡在他的计划上；他要选择，是忠于他的祖国留在莫斯科，还是讨厌地离开它来美国；是来美国过一种粗俗的生活（因为他不喜欢我们对穷人的态度，而且我们缺少对每个人的医疗关怀），还是留在莫斯科痛苦地在无能官僚的淫威下生活；他一方面想享有美国的自由和财富，另一方面却害怕克格勃对家庭、朋友甚至他本人（假如他“叛变”的话）的迫害。最后他说，不去了。他向我推荐了格拉斯哥大学的德雷维尔（Ronald Drever）。

我咨询的其他人也热情推荐德雷维尔。他和布拉金斯基一样，极富想象力和创造力，而且有坚强的意志——这些都是计划成功的基本素质。加州理工学院教授委员会和行政部门尽可能地收集了有关德雷维尔和其他候选者的材料，最后选定了他，请他来学院启动计划。他跟布拉金斯基一样很痛苦，但还是答应了。于是我们离开了莫斯科。

我在提出计划时也想像韦伯和布拉金斯基那样在加州集中力量搞棒探测器。回想起来，幸运的是德雷维尔坚持走完全不同的路。他在格拉斯哥与棒探测器打了5年交道，知道它们有什么局限。他认为，更有希望的是干涉仪式的引力波探测器（简称干涉仪——当然完全不同于第9章的射电干涉仪）。

用于引力波探测的干涉仪最早是由布拉金斯基的两个俄罗斯朋友吉尔增什坦（Mikhail Gertsenshtein）和普斯托瓦特

（V.I.Pustovoit）在1962年想到的，他们提出了它的原始形式。1964年，韦伯也独立想到了。外斯（Rainer　Weiss）在不知道这些早期思想的情况下，在1969年设计了更成熟的干涉仪探测器。接着，他和他在麻省理工学院（MIT）的小组继续设计并在1970年初制造了一台。这时，福瓦德（Robort　Forward）和他在加利福尼亚马里布休斯研究实验室的同行们也在做同样的事情。[11]他们的探测器是第一次运行成功的。到70年代后期，这些干涉仪探测器已经成为探测棒的重要替代者，德雷维尔也为它们的设计贡献了自己的聪明和技巧。[12]

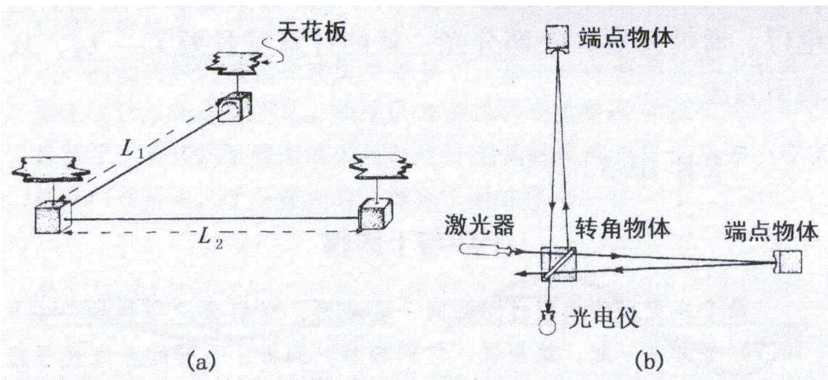



图10.6　激光干涉仪式引力波探测器。这种仪器很像迈克尔逊和莫雷1887年用来探寻地球在以太中运动的那种干涉仪（第1章）。详细解释见正文。

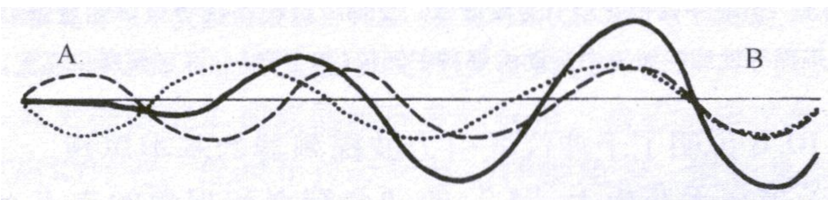
图10.6说明了干涉仪式引力波探测器的基本思想。三块物体由绳子吊在天花板上“L”的两个端点和拐角的支点上（图10.6（a））。当引力波的第一个波峰从屋顶或地板进入实验室时，潮汐力将沿“L”的一臂把两个物体分开，而沿另一臂把两个物体拉近。结果，第一臂的长度（即臂上两个物体间的距离）$L_1$将增大，而第二臂的长度$L_2$将减小。当第一波峰过去，波谷到来时，伸长和缩短的方向会发生改变：$L_1$将减小，$L_2$将增大。通过测量臂长差$L_1$-$L_2$，我们就能发现引力波。

臂长差$L_1$-$L_2$是通过干涉仪（图10.5（b），卡片10.3）监测的。让一束激光照在转角物体的光束分离器（分光镜）上，则光束的一半将被反射，另一半透过去，这样一束光就分成了两束。这两束光将沿

着干涉仪的两臂达到两个端点，然后被端点物体上的镜面反射回到分离器，每一束光仍被分成两束。这样，每束光都有一部分与另一束光的一部分结合，回到激光器；另外两个部分结合到达光电仪。如果没有引力波，则来自两臂的光的干涉结果是，干涉后的光都回到了激光器，不会有到达光电仪的光。如果引力波稍微改变了$L_1$-$L_2$，则两臂的光束将经过稍微不同的距离，从而干涉也会略有不同——有少量联合的光会进入光电仪。通过测量这一部分光，就能计算臂长差$L_1$-$L_2$，从而发现引力波。

卡片10.3　干涉与干涉仪

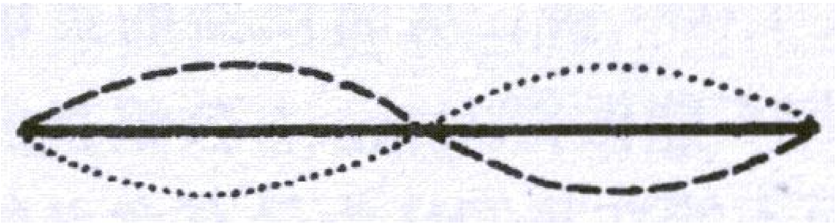
两个或更多的波经过空间同一区域时，它们会“线性地”（卡片10.1）叠加在一起，就是说，它们相加。例如，下面的点线波与虚线波的叠加产生实线波：



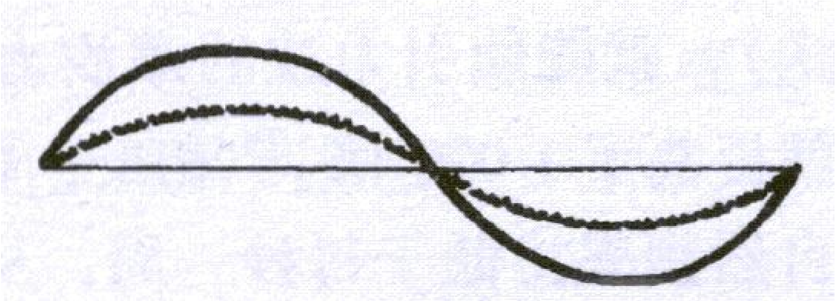

注意A点那样的位置，一个波的波谷（点线）叠加到另一个波的波峰（虚线）上，波就抵消或至少部分抵消了，结果波消失或者减弱了（实线）。而在B那样的地方，两个谷或者两个峰相叠加，波彼此增强了。我们说这是波在相互干涉，在第一种情况下相互破坏，第二种情况下相互加强。这样的叠加和干涉可以发生在所有类型的波——水波、电波、光波、引力波——而干涉正是射电干涉仪（第9章）和干涉仪式引力波探测器运行的关键。

在图10.6（b）的干涉仪探测器中，分光镜将来自一臂的光束的一半叠加到来自另一臂的一半上，传到激光器；又将另外的两半光束叠加起来送到光电仪。如果没有引力波或其他力移动物体和镜面，则

叠加的光波有如下的形式：



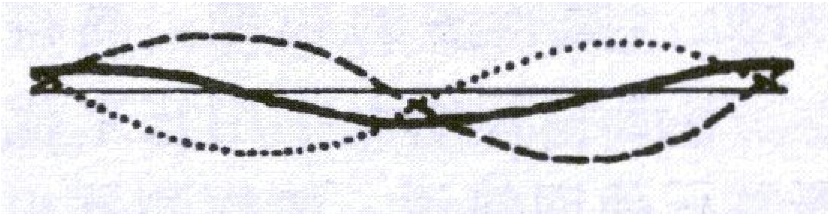
向着光电仪



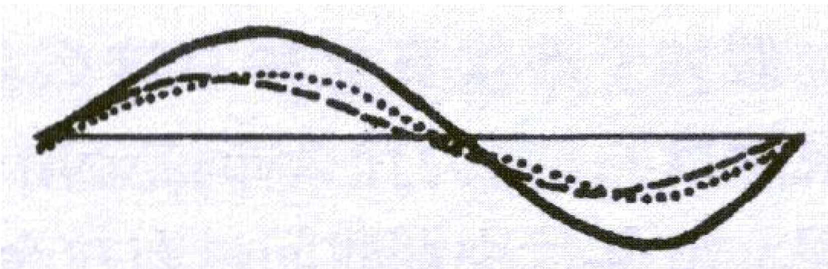
向着激光器

图中的虚线代表来自第一臂的波，点线是来自第二臂的波，实线是叠加的结果。

到光电仪的波完全被干涉破坏了，所以叠加结果是零，这意味着光电仪什么光也看不见。如果引力波或其他力将某臂拉长而将另一臂缩短了，那么从长臂来的光束到达分光镜的时间相对于从另一臂来的将有一点延迟，于是叠加的波就像下面这样：



向着光电仪



向着激光器

沿光电仪方向的光不再被干涉完全破坏，还能收到一些。收到的量正比于臂长的差$L_1$-$L_2$，而这个差正比于引力波信号的强度。

比较棒探测器与干涉仪是很有意义的。棒探测器通过一根实心圆柱的振动来监测引力波的潮汐力；干涉仪探测器通过悬挂物体的相对运动来监测潮汐力。

棒探测器以电传感器（如被棒挤压的压电性晶体）来监测由波引起的棒的振动。干涉仪探测器通过干涉的光束来监测波动引起的物体运动。

棒只对很窄频率范围内的引力波才会产生共振响应，所以解译交响的引力波需要一个由许多棒构成的“木琴”。干涉仪的物体对所有频率高于每秒一周的波都会前后摆动地响应，[2]因此干涉仪有很宽的频率范围，三四个这样的干涉仪就足够完全解读引力波的交响了。

将干涉仪的臂做得比棒长1 000倍（也就是几千米，而不是几米），就能使引力波的潮汐力大1 000倍，从而仪器的灵敏度也就提高了1000倍。[3]相反，棒却不能做得太长。1公里长的棒的自然频率将低于每秒一周，从而不能在我们认为最有意义的频率范围内工作。而且，因为频率这么低，必须把棒发射到太空去，将它与地面的振动和地球大气的重力波动隔绝开来。把这样一个棒放到太空是很荒唐的，而且不知要花多少钱。

因为干涉仪比棒长了1 000倍，它对测量过程产生的反冲作用的“免疫力”也提高了1000倍。这儿的“免疫力”说的是，这种干涉仪不需要靠什么量子无破坏探测器（那也是很难做的）来克服那些反冲作用。相反，棒只有在使用了量子无破坏技术时才可能探测到希望的波。

假如干涉仪真比棒有这么多好处（宽得多的频带，高得多的灵敏度），那为什么布拉金斯基、韦伯和其他人不用它来代替棒呢？70年代中期我问过布拉金斯基。他回答说，棒探测器很简单，而干涉仪却复杂得吓人。像他在莫斯科的那个精干小组完全有可能造一个能很好运行、足以发现引力波的棒探测器。然而，要制造、改进并成功运行一组干涉仪探测器，需要大量的人和大笔的钱——而且，即使有了这么些人和这么多钱，布拉金斯基也怀疑那么复杂的探测器是否能成功。

10年后，越来越多的证据令人痛苦地表明，棒探测器很难达到$10^{-21}$的灵敏度。这时，布拉金斯基来访问加州理工学院，德雷维尔小组用干涉仪取得的进展感动了他。他承认，干涉仪最终会成功的。但为了成功而耗费大量的研究力量和金钱并不令他喜欢。所以，一回到莫斯科，他就将他的小组的大部分力量转到远离引力波探测的方向去了。[13]（世界上还有些地方在继续发展棒探测器，那是幸运的；它比干涉仪便宜多了，现在也更灵敏；它们最终可能在高频率引力波方面发挥特别重要的作用。）

那么，干涉仪探测器复杂在哪些地方呢？毕竟，图10.6描述的基本思想看起来是非常简单的。

事实上，图10.6是超级简化了的，它忽略了数不清的陷阱。为避免这些陷阱需要很多技巧，这就使干涉仪变得复杂了。举例说，激光束必须精确地指向一定的方向，精确地具有一定的形状和波长，这样才能完全适合于干涉仪；而且，它的波长和强度不能起伏波动。光束一分为二后，两束光分别在两臂上来回反射，不是图10.6所画的一次，而应该是多次，这样才能提高它们对摆动物体的运动的敏感性；多次反射后，它们还必须正好回到分光镜。每一个悬挂的物体必须在不断的控制下，使镜面精确指向同一个方向，不会因为地板的振动而摇摆，而这样的控制还不能掩盖引力波引起的物体摆动。为了在所有这些方面以及其他许许多多方面达到完善，需要不断地监测干涉仪的许多不同部位和它的光束，还要不断利用反馈的力来保持完好的状态。

从下面的照片（图10.7）你可能会得到一些复杂的印象。那是德雷维尔小组在加州理工学院建造的一个40米长的原型干涉仪探测器——它比所需要的原大几千米长的干涉仪简单得多。

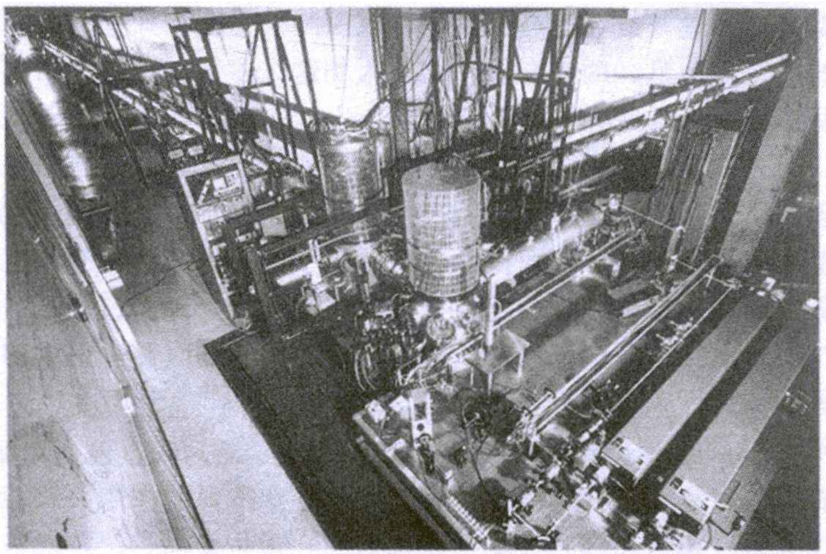


图10.7　加州理工学院的40米原型干涉仪引力波探测器（约1989年）。放在前面的桌子和笼子里的真空室装着激光器和让激光进入干涉仪的设备。中心物体放在第二个笼子的真空室里——可约隐看到它上面吊着的绳子。两个端点物体沿走廊在40米以外。两臂的光束在两根真空管中较大的一根内往返，真空管是用来增大臂长的。[加利福尼亚理工学院LIGO计划提供。]

80年代初期，有四个实验物理学家小组在努力发展干涉仪探测器的工具和技术：德雷维尔的加州小组，他在格拉斯哥创建的小组（现在由霍克（James Hough）领导），外斯的MIT小组和比林（Hans Billing）在德国慕尼黑马克斯·普朗克研究所建立的小组。这些研究队伍都小而精，或多或少都在独立进行工作，4用各自的方法来设计干涉仪探测器。组内的每个科学家都可以自由地提出他的新观点，可以照自己的意愿去发展它们，而且时间也是想多长都行。这是非常轻松的科学合作形式，是有创造力的科学家所喜欢的，他们也正是在这种文化环境下成熟起来的；这也是布拉金斯基渴望的，像我这样的孤独者更能在其中感到快乐。但是，它并不适宜于复杂科学仪器（如我们需要的几公里长的干涉仪）的设计、建造、改进和运行。

为了详细设计这样一个干涉仪的许多复杂部件，为了让所有部件能组装起来正常运行，为了把费用控制在计划内并在有限时间里完成

干涉仪，需要一种不同的工作模式：一种密切协作的模式，每个组的各小组要集中到一个确定好的目标上来，每个负责人要决定该做什么，谁来做，什么时候做。

从自由独立走向密切协作是很痛苦的。生物学家们在为人类染色体排序时曾痛苦地经历过这样的历程。[5]从1984年起，我们引力波物理学家也上路了，也一样少不了痛苦和悲伤。然而我相信，总有一天，引力波的发现和解译所带来的激动、快乐和科学回报，将把这些痛苦和悲伤从我们的记忆中抹去。

走上这条痛苦之路遇到的第一个大转折是1984年加州和麻省的两个小组被迫合并——那时每组有8个人。美国国家科学基金会（NSF）的伊萨克逊（Richard Issacson）为了纳税人的财政支持，强迫这两个学院的科学家走到一起来联合发展干涉仪。德雷维尔坚决反对，而外斯看到走不脱了，只好答应。两个人像被迫结婚的新人，发了誓，我成了他们的和事佬。如果两人分道扬镳，我有责任把他们拉回来。这是脆弱的婚姻，大家都没有一点儿感情。不过，我们还是慢慢开始一起工作了。

第二个大转折出现在1986年。一个由知名物理学家组成的专家委员会——包括我们需要的所有技术方面的专家和科学大项目的组织管理专家——来我们这儿一个星期，检查了我们的进展和计划，然后向NSF报告。我们的成绩和计划都获得了高度评价；我们成功发现并解译引力波的前景也被认为是大有希望的。但是，给NSF的报告说，我们的组织很糟，还是原来那种松散自由的结合形式，照这样是永远也不可能成功的。委员会认为，应该用一个领导者来代替德雷维尔-外斯-索恩的三足鼎立——他能将单个的人组织成一个紧密团结的能干的小组，能组织项目并能在每一个紧要关头做出果断明智的决定。

压力又来了。NSF的伊萨克逊告诉我们，如果想项目继续下去，就必须找那样一个人来，像足球队员跟一个伟大的教练和乐队跟一个

伟大的指挥那样跟他一起工作。

在寻找中，我们幸运地发现了福格特（Robbie Vogt）。

福格特是一个才华横溢、意志坚强的实验物理学家，曾领导宇宙飞船科学仪器的制造和试验，领导过巨型毫米波天文干涉仪的研制，还组织过美国宇航局喷气实验室[6]（大多数美国行星探测计划都是在这儿执行的）的科学研究——后来，他成为加州理工学院的教务长。虽然福格特是一个特别能干的教务长，但他与院长戈尔德贝格（Marvin Goldberger）在如何领导管理学院问题上有过激烈争吵——吵了几年，戈尔德贝格便将他解聘了。福格特的个性不适合在别人手下工作，特别是当他与他们的观点有重大分歧的时候；不过他会是一个很好的头儿。他就是我们需要的那个领导者，那个指挥，那个教练。如果说有人能让我们紧密团结起来，那就是他了。

“跟罗比工作是很痛苦的。”他原来毫米波小组的人告诉我们，“你们将留下创伤，不过那也值。你们的计划会成功的。”

德雷维尔、外斯、我和其他一些人同福格特谈了几个月，请他来做我们的领导。最后他答应了。真像说的那样，我们原来的加州-麻省小组终于在6年后被打破了，但新的小组更紧密、更有力、更有效，很快壮大到约50位科学家和工程师，都是成功所需要的人。然而，成功不是靠我们一家就够了。在福格特计划下，别的科学家也为我们的中心研究作出了重要贡献。[7]他们松散地与我们联系，还能保持我们留下的独立和自由。



1991年下半年LIGO计划加州-麻省小组的部分科学家。左：组内的部分加州成员，左上起反时针方向：Aaron Gillespie,Fred Raab,Maggie Taylor,Seiji Kawamura,福格特，德雷维尔，Lisa Sievers,Alex Abramovici,Bob Spero,Mike Zucker。右：组内部分麻省成员，左上起反时针方向：Joe Kovalik,Yaron Hefetz,Nergis Mavalvala,外斯，David Schumaker,Joe Giaime。[左图由Ken Rogers/Black Star提供；右图由Erik L.Simmons提供。]

在我们的努力中，成功的关键是建立并启用一套全国性的科学装置，叫做激光干涉仪引力波天文台，或LIGO。[14]LIGO由“L”型真空系统构成，一个在华盛顿汉福德附近，另一个在路易斯安娜利文斯顿附近。物理学家要在这儿开发和运行一系列的不断改进的干涉仪，见图10.8。

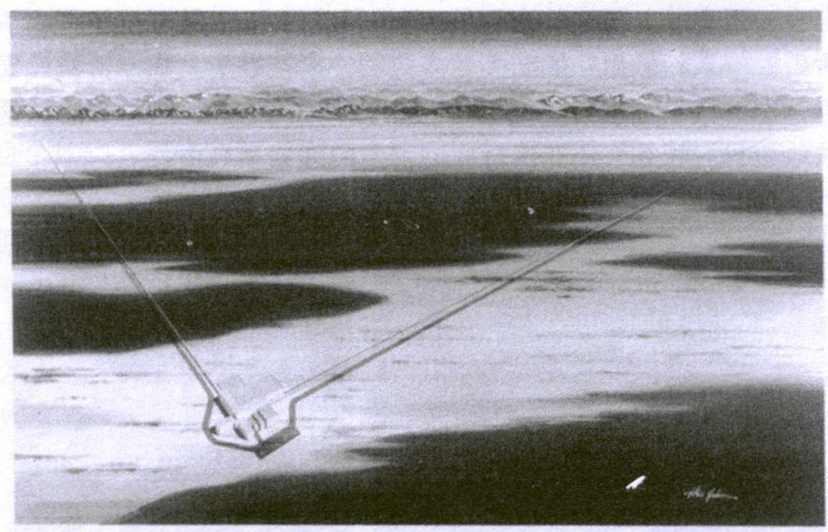


图10.8　艺术家心目中的LIGO“L”型真空系统和在华盛顿汉福德附近“L”中心的实验基地。[加利福尼亚理工学院LIGO计划提供。]

为什么要两个实验基地，而不是一个呢？因为地球上的引力波探测器总会将噪声误会成引力波的爆发。例如，悬挂物体的绳子会无故轻微摆动，像引力波潮汐力那样摇动物体。然而，这样的噪声几乎不可能同时发生在两个远离的独立探测器。因此，为了保证明显的信号来自引力波而不是噪声，必须确认它在两个探测器上都出现。一个探测器是不可能发现并监测引力波的。

虽然两个探测器就足以探测到引力波了，但我们实际需要三个，四个更好。这些远远分开的探测器可以完全解译交响的引力波，也就是将波所携带的信息完全析取出来。一个法国-意大利联合小组将在意大利比萨附近建立第三个基地，名叫VIRGO。8VIRGO和LIGO将形成一个全息的国际探测网。英国、德国、日本和澳大利亚正在筹资准备为这一网络建立另外的基地。

为一种谁也不曾见过的波建立那么庞大的网络，似乎胆子也太大了。实际上那完全不是胆大。引力波已经被天文学观测证实存在了，普林斯顿大学的泰勒（Joseph Taylor）和赫塞（Russel Hulse）为此

获得1993年度诺贝尔奖金。他们用射电望远镜发现了两颗中子星，其中一颗为脉冲星，它们每8小时互相绕着旋转一周。通过极精确的射电测量，他们证明两颗星以爱因斯坦定律所预言的速率（每年十亿分之二点七）螺旋式地靠近，原因是它们向宇宙中发出的引力波所持续产生的反冲作用。除了引力波的小小反冲作用，没有别的原因能解释这两颗星的螺旋靠近。

21世纪初的引力波天文学会是什么样子呢？我们可以想象下面的景象：

2007年，八个几公里长的干涉仪在全天候地运行，扫描天空，寻找到来的引力波。这八个干涉仪，两个运行在意大利比萨的真空装置里，两个在美国东南路易斯安娜州的利文斯顿，两个在美国西北华盛顿的汉福德，还有两个在日本。每个地方的两个干涉仪，有一个是“服劳役的”机器，监测振荡频率在每秒10到1 000周范围内的波；另一个才新近研制安装，是先进的“做研究的”干涉仪，瞄准每秒1 000到3 000周的振荡。

一列引力波从宇宙遥远的源头掠过太空来到太阳系。一个波峰首先落在日本的探测器上，然后穿过地球到达华盛顿，接着到路易斯安娜，最后到达意大利。大约1分钟，波谷跟着波峰来，波峰又跟着波谷来。每个探测器的悬挂物体轻轻转动，干扰了激光束，从而也干扰了进入探测器光电二极管的光。八个光电二极管的输出信号通过卫星网传到中心计算机，计算机提醒科学家，另一列1分钟的引力波已经来到地球，是本周的第三波。计算机结合八个探测器的结果，要完成四件事情：引力波爆发源在天空的最佳位置估计；位置估计的误差区间；两个波形——即两条振荡曲线，类似于检测声波时在示波器上看到的振荡曲线。波源的历史就藏在这些波形曲线里（图10.9）。

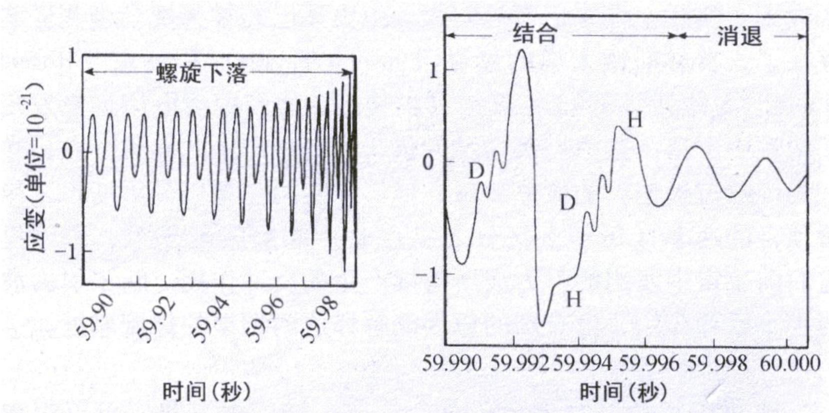



图10.9　黑洞结合所产生的两个波形之一。竖直方向是以$10^{-21}$为单位的应变；水平方向是以秒为单位的时间。第一幅图只画了波的螺旋下落过程的最后0.1秒；波形在前1分钟很简单，只是振幅与频率逐渐增大。第二幅图夸张地画了最后0.01秒的波形。1993年，根据爱因斯坦场方程的解，已经很好地认识了波形的螺旋和衰落（消退）阶段。结合阶段还完全不了解（图中的曲线是我个人的想象）；未来的超大计算机将对它进行计算模拟。在正文中，我假定这些模拟在21世纪初已经成功了。

之所以有两个波形，原因是引力波有两个极化。如果波垂直通过干涉仪，则一个极化描述了沿东西向和南北向振荡的潮汐力；另一个极化描述了在东北-西南方向和西北-东南方向振荡的潮汐力。因为每个探测器都有自己的定向，所以它们收到的是这两个极化的某种组合；计算机要从八个探测结果中重新找出那两个极化波形。

然后，计算机将得到的波形与一个大波谱表里的波形进行对比，这很像鸟类观察者通过与图谱的比较来识别一只鸟。经过5年对来自碰撞、结合的黑洞、中子星、旋转中子星（脉冲星）和超新星爆发的引力波的监测经验和计算机对波源的模拟，这样的波谱表已经做出来了。引力波的爆发是很好确认的（另外一些波，如来自超新星的，就困难得多）。波形确定无疑地显示了两个黑洞结合的惟一信号，它包括如下三段：

·1分钟长的第一段（图10.9只画了最后0.1秒）具有振幅和频率都逐渐增大的振荡应变，正是我们预料的来自双星轨道上两个螺旋靠近

的天体的波形。波的大小交错变化说明轨道像椭圆，不是正圆。

·0.01秒长的中间段几乎完全符合超大计算机最近（21世纪初）对两个黑洞结合过程模拟的预言。根据模拟，标记“H”的峰表示两个黑洞的视界的接触与融合。然而，标记“D”的两个摆动却是新设计的那个“做研究的”干涉仪的第一个新发现，那些老的“服劳役的”干涉仪从来没能探测到这些摆动，因为它们频率太高了；而且它们在超大计算机模拟中也从没出现过。这是理论家需要解释的难题。也许它们第一次提供了某些线索，能帮助我们认识碰撞黑洞的时空曲率非线性振动中我们不曾料想的奇异行为。为这种景象所迷惑的理论家该回到他们的模拟中去寻找这对摆动的信号。

·0.03秒长的第三段（图10.9只画了它的开头）由频率固定而振幅衰减的振荡构成。我们预料变形的黑洞在为摆脱形变而脉动时就会产生这样的波动；就是说，这样的波像落幕的铃响，慢慢衰落下去。脉动的是两个哑铃型的突起，它们绕着黑洞赤道一圈圈地旋转，随着能量逐渐被曲率波带走，它们也将消失（图10.2上）。

根据这些波形的细节，计算机不但能解析黑洞碰撞、结合和衰落的历史，还能计算初始黑洞和终结黑洞的质量和旋转速度。每个初始黑洞有25个太阳那么重，旋转很慢；终结黑洞有46个太阳那么重，以最大允许转速的97%旋转。与4个太阳（2×25-46＝4）质量相当的能量转化为曲率波，随波飘散了。初始黑洞的总表面积是136　000平方公里，终结黑洞的表面积更大，144000平方公里，这是黑洞力学的第二定律要求的（第12章）。波形还揭示了黑洞距地球的距离：10亿光年，这个结果大约有20%的精度。波形还告诉我们，以前的视线近似垂直于轨道平面，现在我们从旋转黑洞的北极看下去，两者比较说明，黑洞的轨道有30%的偏心率（长圆形的）。

根据波峰到达日本、华盛顿、路易斯安娜和意大利的时间，计算机确定了黑洞在天空的位置。因为波先到日本，所以它多少在日本的

头上，而在美洲和欧洲脚下。详细分析到达时间，可以为波源确定一个误差区间为1度的最佳猜测位置。如果黑洞更小，波形振荡会更快，误差区间将更小，但对这些大黑洞，探测网只能做到1度的水平。再过10年，在月亮上运行干涉仪探测器时，误差区间将在某些方面减小100倍。

因为黑洞轨道被拉长了，计算机判断两个黑洞从互相捕获到绕对方旋转的轨道到结合和发射引力波，只有几个小时。（如果它们在轨道上旋转的时间超过几个小时，离开它们的引力波的反冲作用将使轨道成为圆形的。）那么快的捕获说明黑洞可能在某星系中心的一个致密的由黑洞和大质量恒星组成的集团之中。

于是，计算机接着检查光学星系、射电星系和X射线星系表，寻找那些距地球8亿到12亿光年、在1度误差区间内的有特殊核的星系。它为天文学家找到了40个候选者。在接下来的几年里，射电的、毫米波的、红外的、光学的、紫外的、X射线以及γ射线的望远镜将对这40个候选者进行详细的研究。我们会逐步认识到，在某一个候选星系的核心聚集着大量的气体和恒星，当我们现在看到的光离开它时，那里正在展开一幕百万年的剧烈演化——巨黑洞将在演化中诞生，类星体也将随演化而形成。感谢引力波的爆发，它为这个特别的星系带来了意义，天文学家现在可以去揭示巨黑洞是怎样诞生的了。

[1]　这些差别，它们的结果以及所预料的来自不同天体物理源的波的具体情况，已经有很多科学家阐述过了，他们包括（当然还有别人）：巴黎的Thibault Damour，莫斯科的Leonid Grishchuk，京都的Takashi Nakamura，威尔士的Bernard Schutz，纽约绮色佳的Stuart Shapiro，圣路易的Clifford Will，还有我。

[2]　如果频率低于每秒一周，悬挂物体的绳子会阻止它们响应那些波，物体也就不能摆动了。

[3]　实际上，具体情况比这复杂得多，灵敏度的提高也远不是像这几句话说的那样容易实现；不过，这里讲的大体上还是正确的。

[4]　不过，格拉斯哥和加州的小组通过德雷维尔而有着密切联系。

[5]　人类基因组计划1990年在美国启动，英、日、法、德和中国科学家先后加盟，历经10年，在2000年6月26日完成了人类基因组草图绘制工作，测定了DNA中90%以上的碱基序列。这是比“曼哈顿”原子弹计划、“阿波罗”登月计划影响更为深远的科学计划。——译者

[6]　这个实验室就在帕萨迪纳的西北，是美国宇航局委托加州理工学院管理的。——译者

[7]　以1993年为例，包括莫斯科的布拉金斯基小组，斯坦福大学Bob Byers领导的小组，科罗拉多大学的Jim Faller小组，锡拉丘兹大学的Deter Saulson小组，以及西北大学的Sam Finn小组。

[8]　名字来自室女（Virgo）星系团，有可能探测到它的引力波。

# 第11章　实在是什么？

时空，

在星期天弯曲，在星期一平直；

视界，

在星期天是真空，在星期一是电荷；

而实验，

在星期天和星期一都是一样的

时空真是弯曲的吗？能不能这样想，时空本来是平直的，但我们用来测量它的钟和尺——我们认为理想的（什么是理想的，请看卡片11.1）钟和尺，实际上却是“橡皮的”？当我们从一点走到另一点改变它们方向的时候，即使最完美的钟也可能慢或者快，即使最完美的尺也可能缩或者长，不是吗？我们的钟和尺的这些变化，是不是会让一个平直的时空显得弯曲呢？

是的，完全可以那么想。

图11.1举了一个具体的例子：测量非旋转黑洞的周长和半径。左边是黑洞弯曲空间的嵌入图，在这个图中，空间是弯曲的，因为我们已经假定我们的尺子不是橡皮的，不论把它放在什么地方，让它指向什么方向，它都会保持自己的长度——距离就是照这样定义的。尺子测量的黑洞视界的周长是100千米。洞外还画了一个两倍周长，即200千米的圆，从视界到这个圆的径向距离也用理想的尺子来量，结

果是37千米。假如空间是平的，那么径向距离应该是外圆的半径，200/2π千米，减去视界的半径，100/2π千米；也就是，200/2π-100/2π＝16千米（近似）。为了满足那个大得多的径向距离37千米，表面必然表现为像图中那样弯曲的喇叭形。

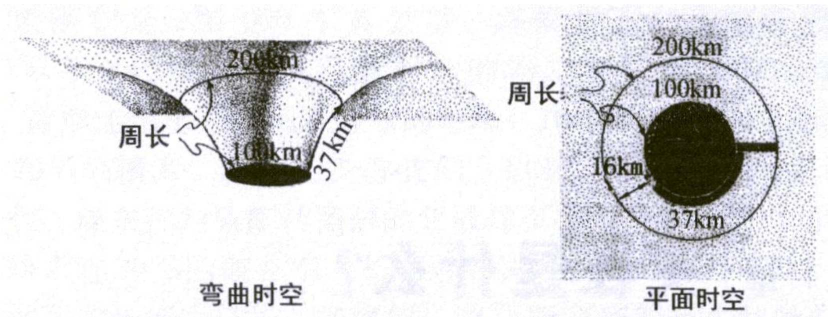



图11.1　两种观点下黑洞附近的长度测量。左：认为时空真是弯曲的，理想的尺子精确测量时空的长度。右：认为时空实际是平直的，而理想的尺子是橡皮的，它能精确测量真实的平直时空的长度。然而，在指向径向时，它会发生收缩，离黑洞越近，收缩量越大，因此它测得的径向长度比真实的大（在图示的情况下，它测的是37千米，而不是真正的16千米）。

卡片11.1　理想的钟与尺

在这本书里，我说“理想的钟”和“理想的尺”，全世界最好的钟表和量尺制造者都明白它的意思：理想需要通过与原子和分子的行为进行对比来认定。

更具体地说，理想的钟在与原子或分子振荡对比时，必须均匀地“嘀嗒”。世界上最好的原子钟就是设计来做这个的。因为原子和分子的振荡是由我以前说的“时间流的速率”决定的，这就意味着，理想的钟测量爱因斯坦弯曲时空的“时间”部分。

理想尺子的刻度与原子或分子发出的光的波长比，必须有均匀而标准的间隔。例如，相对于氢分子发出的21厘米波长的光的均匀间隔。这相当于要求，在某一固定标准温度（如0℃）下，尺子在两个刻度之间总是包含同样固定数目的原子；这也反过来确保了理想尺子

测量爱因斯坦弯曲时空的“空间”长度。

这一章引进了“真”时间和“真”长度的概念，它们并不一定是理想的钟和尺所测量的时间和长度，也就是说，不一定是原子和分子标准的时间和长度，也不一定是嵌入爱因斯坦弯曲时空的时间和长度。

假如黑洞周围的空间本是平直的，而我们因理想尺子是橡皮的而误认为空间是弯曲的，那么真正的空间几何一定像图11.1右图那样，视界和圆之间的真实距离应该是平直的欧几里得几何定律所要求的16千米。然而，广义相对论认为，我们的理想尺子没有测量这一真实距离。拿一把尺子沿着黑洞周长放在视界上（如图11.1右边的带刻度的弯黑宽带），尺子像这样沿着周长方向，的确在测量真实距离。从尺子切出37千米长的一段，它覆盖黑洞周边的37%。然后，将尺子转到半径方向（图上带刻度的直黑宽带）。转向时，广义相对论要求它发生收缩，指到径向时，它的真实长度一定已经收缩到16千米了，刚好从视界到达外面的圆。但是，收缩表面的尺度还是认为它的长是37千米，从而视界与圆之间的距离是37千米。这样，像爱因斯坦那样不知道尺子是橡皮的人会相信这个不准确的测量，认为空间是弯曲的。而像你我这样认识了橡皮特性的人却知道，尺子收缩了，空间还是平直的。

什么东西能让尺子在改变方向时发生收缩呢？当然是引力。在图11.1右边的平直空间里，存在着决定一切事物（包括基本粒子、原子核、原子、分子等）大小的引力场，它迫使所有事物在径向上收缩。离黑洞越近，收缩量越大，离黑洞越远，收缩量越小，因为决定收缩的引力场是黑洞产生的，它的影响随离开黑洞的距离而减弱。

决定收缩的引力场还有其他效应。如果光子或其他粒子飞过黑洞，引力场作用将使它的轨迹发生偏转，在黑洞周围，轨迹是弯的；在黑洞真实的平直时空几何中测量，它是曲线。但像爱因斯坦那样看重他们的橡皮尺钟测量的人，认为光子是在弯曲时空里沿直线运动。

哪个是真正的事实呢？时空是像上面说的那样平直呢，或者还真是弯曲的？对物理学家（如我）来说，这个问题很无聊，因为它没有物理意义。无论弯曲的还是平直的，两种时空观对任何理想尺钟所进行的测量都做出完全相同的预言，而且我们会看到，它对任何类型的物理仪器所进行的测量，也做出相同的预言。例如，两个观点都同意，图11.1中视界与图之间的径向距离正是理想尺子所测量的37千米。他们的争论在于这个测量距离是否“真实”，但这属于哲学争论，不是物理学的。因为两者在任何实验的结果上都一致，所以它们在物理学上是等价的。至于哪个观点告诉了“真正的事实”，是与实验无关的，那是哲学家而不是物理学家要讨论的问题。另外，物理学家在推导广义相对论预言时，可以而且确实交换地运用这两个观点。

库恩（Thomas　Kuhn）的规范概念，[1]很好地描述了理论物理学家工作的智力活动。1949年，库恩在哈佛大学获得物理学博士学位，后来成为著名的科学哲学家。在1962年的《科学革命的结构》一书里，[1]他提出了规范的概念——那是我读过的最有见识的一本书。

一个规范就是科学家群体在研究某个问题和与别人交流研究结果时所用的一整套工具。在广义相对论上，弯曲时空观是一个规范，平直时空观是另一个规范。每个规范包括三个基本因素：一组数学化的物理学定律；一组供我们洞察定律和帮我们与人交流的图像（头脑里的、口头上的和画在纸上的）；一组典型事例——即过去的计算和已经解决的问题，可以是教科书上的，也可以是科学论文里的。它们都是相对论专家们认为做得很好、很有意义的，我们拿来作为未来计算的样本。

弯曲时空规范以三组已经建立的数学化定律为基础：爱因斯坦场方程，它描述物质如何产生时空曲率；告诉我们理想尺钟测量爱因斯坦弯曲时空的长度和时间的定律；告诉我们物质和场如何在弯曲时空中运动，例如，自由运动的物体沿直线（测地线）运动的定律。平直时空规范也以三组定律为基础：描述平直时空中的物质如何产生引力

场的定律；描述场如何决定理想尺寸的收缩和理想的时钟流如何膨胀的定律；描述引力场如何决定粒子和场在平直时空中运动的定律。

弯曲时空规范的图像包括本书画过的嵌入图（如图11.1左边）和对黑洞周围时空曲率的语言描述（例如，“旋转黑洞周围的龙卷风式的旋涡”）。平直时空规范里的图像包括图11.1的右边，一把在从周长方向转到半径方向发生收缩的尺子，以及对“决定尺子收缩的引力场”的语言描述。

弯曲时空规范的典型事例包括能在大多数相对论教科书里看到的计算，可以用这些计算导出爱因斯坦场方程的史瓦西解，还包括伊斯雷尔、卡特尔和霍金等人推演黑洞“无毛”的计算。平直时空规范的典型事例包括教科书里关于黑洞和其他物体在捕获引力波后质量如何变化的计算和韦尔（Clifford Will）、达莫尔（Thibault Damour）等人关于相互围绕转动的中子星如何产生引力波（收缩产生的场的波动）的计算。

在我做研究时，规范的每一部分——定律、图像和范例——对我的思想过程都是很重要的。图像（头脑中的、口头上的和纸上的）像指南针，为我带来对宇宙行为的直觉；凭它们和一些数学的草算，我可能找到一些有趣的想法。如果找到了值得追求的东西（如第7章的环猜想），我会在规范的数学化物理定律的基础上做进一步计算，证明或否决它。详细计算可向规范中的范例学习，它们将告诉我，可靠的结果需要多高的精度。（如果精度太低，结果可能是错的；如果精度太高，计算将浪费不少时间。）范例还会告诉我，哪样的计算能帮我通过数学符号的泥潭达到我的目标。图像也能指导计算，帮我发现捷径，避开死胡同。假如计算证实了我的新想法，或者至少说明它似乎是合理的，我会通过图像和计算与相对论专家们交流，也用图像——口头的和书面的，与其他人交流，比如这本书的读者。

平直时空规范的物理定律可以从数学上根据弯曲时空规范的定律

推导出来，反过来也行。这就是说，两组定律是同一物理现象的不同数学表示，有点像用0.001和1/1 000来表示同一个数。不过，定律的数学公式在两种表示中看起来是很不一样的，相应于两组定律的图像和范例也大不相同。

举一个例子。在弯曲时空规范里，爱因斯坦场方程在口头上可以说“质量产生时空曲率”。用平直时空规范的语言，场方程被说成“质量产生决定尺子收缩和时钟膨胀的引力场”。虽然爱因斯坦场方程的这两个说法在数学上等价，在语言上却大不相同。

在相对论研究中，学会两种规范是极有好处的。有些问题在弯曲时空规范里容易很快解决，另一些问题则需要平直的规范。黑洞问题（如黑洞无毛的发现）最适合用弯曲时空的技巧；引力波问题（如计算两颗中子星相互围绕转动时发出的波）则适合用平直时空的技巧。理论物理学家在成熟中会逐渐觉悟在哪种情形该用哪种规范，他们知道根据需要将问题从一个规范转移到另一个规范来考虑。星期天他们考虑黑洞时，可能认为时空是弯曲的，而在星期一他们考虑引力波时，可能又认为时空是平直的。这样的思路转移，我们在看埃舍尔（M.C.Escher）的画时会同样经历。例如图11.2。[2]

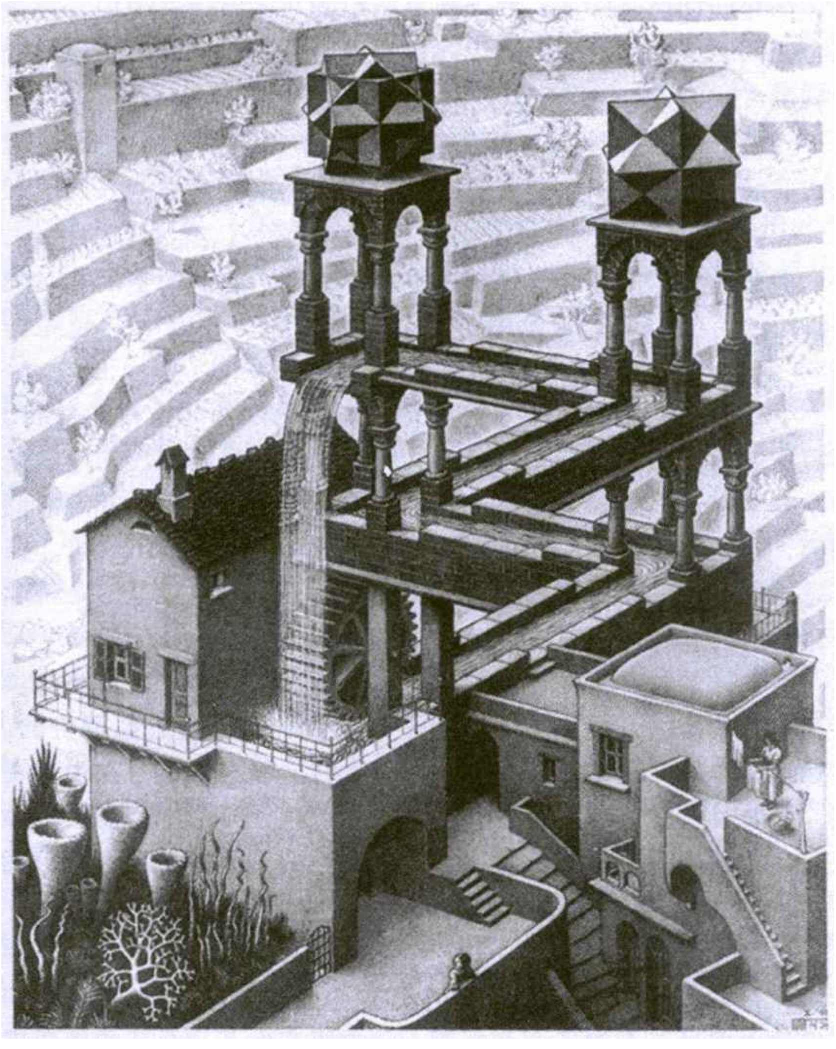


图11.2　埃舍尔的一幅画。先从一点看（例如，从瀑布顶的流水看），然后从另一点看（例如，从瀑布底的流水看），我们会经历一次思路转移，多少有点儿像物理学家从弯曲时空规范转到平直时空规范的经历。[©1961 M.C.Escher Foundation-Baarn-Holland，版权所有。]

因为两个规范的基础定律在数学上是等价的，我们可以确信，在相同物理条件下，两个规范所给的对实验结果的预言将是完全相同的。这样，我们可以在任何给定条件下自由运用最适合的规范。

自由带来力量，[2]这就是为什么物理学家不满足于爱因斯坦的弯

曲时空规范，而还要发展平直时空规范来作为补充。[3]

牛顿的引力描述今天仍然还是一种规范。它认为空间和时间是绝对的，引力是同时作用在两个物体间的一种力（“超距作用”，第1，2章）。

牛顿的引力规范当然不会和爱因斯坦的弯曲时空规范等价，两家所做的实验结果的预言是不同的。库恩说这场理性的斗争是科学的革命，爱因斯坦通过革命提出了他的规范，令他的同事们相信，新规范比牛顿的规范更准确地描述了引力（第2章）。在库恩的这个意义上，物理学家后来提出的平直时空规范不是科学革命，因为它与弯曲时空规范做出的预言是完全一样的。

引力较弱时，牛顿规范的预言与爱因斯坦弯曲时空规范的预言几乎是一样的，相应地，两个规范在数学上也是近似等价的。实际上就是这样：在研究太阳系的引力时，物理学家常在牛顿规范、弯曲时空规范和平直时空规范之间游移，哪个规范满足他的想象，哪个规范显得更具洞察，他们就用哪一个，而且不会出现问题。[3]

在一个研究领域中，新人的思想总是比老手更加开放，70年代就出现过一个例子，一些新人的觉悟产生了一个新的黑洞规范：膜规范。

1971年，普林斯顿大学的学生汉尼（Richard　Hanni）和博士后鲁菲尼（Remo　Ruffini）注意到，黑洞行为多少有些像一个导电球。为理解这种奇特行为，我们想象一个带正电的金属小球，它携带的电场排斥质子而吸引电子。小球的电场可以用类似于磁力线的电力线来刻画。电力线指向场作用在质子上的力的方向（也就是与场作用在电子上的力的方向相反），线密度正比于力的强度。假如小球独立于时空中，它的电力线将径向向外（图11.3（a））。相应地，作用在质子上的电力也沿径向离开小球。又因为力线的密度随离开小球的距离

的平方反比例地减小，所以作用在质子上的电力也随距离的平方而反比例地减弱。

现在，将小球拿近一个金属球（图11.3（b））。球的金属表面带有可以在表面上自由移动的电子和不能移动的带正电的离子。小球的电场将球面上的大量电子吸引到附近，而把多余的离子留在球面各处，换句话说，小球极化了金属球。[4]

1971年，汉尼和鲁菲尼，另外还有普林斯顿大学的瓦尔德（Robert Wald）和普林斯顿高等研究院的科恩（Jeff Cohen）[4]分别独立计算了非旋转黑洞附近带电小球产生的电力线的形状。他们的计算以标准的弯曲时空规范为基础，结果表明时空曲率像图11.3（c）那样使电力线发生形变。汉尼和鲁菲尼注意到，它与图11.3（b）中的电力线是相似的（从图（c）下面看，与图（b）近似相同），这令他们想到，我们可以用与考虑金属球相同的方式来考虑黑洞的视界；就是说，将视界看作一张由正负带电粒子组成的与金属的球面相似的薄膜。通常情况下，膜上的正负电荷粒子数相等，即膜上任何区域都没有净电荷。然而，当小球靠近视界时，多余的负电粒子会移到小球下面的区域，膜上将到处留下多余的正电粒子，视界膜就这样被极化了。最后，小球电荷和视界电荷所产生的总的电力线就像图（c）的样子。

我算相对论的老人了，听到这些事情时我认为很荒唐。广义相对论主张，如果谁落进黑洞，他在视界那儿除了时空曲率外什么也碰不到。他既看不到膜，也看不到带电的粒子。这样，汉尼-鲁菲尼关于小球电力线为什么会偏折的描述就没有现实基础，纯粹是想象。我确信力线弯曲的原因不是别的而只能是时空曲率：力线向下偏向图（C）中的视界，完全是因为潮汐引力在拉它，而不是因为它被视界的某些极化电荷所吸引。视界不可能有任何这样的极化电荷。我确信这一点，然而我错了。

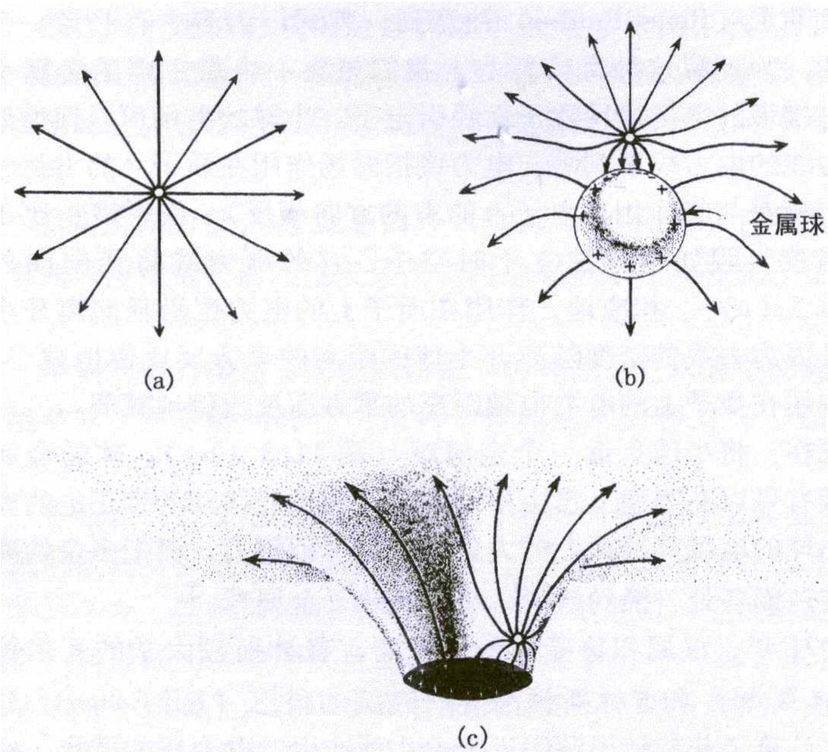



图11.3　（a）独立静止在平直时空的带正电金属小球产生的电力线。（b）平直时空中，小球静止在导电金属球上方时的电力线。小球的电场极化了金属球。（c）小球在黑洞视界上方时的电力线。看起来就像小球电场极化了视界。

5年后，剑桥大学的布兰福德和研究生茨纳耶克发现，磁场可从黑洞中汲取旋转能并用来驱动喷流（即布兰福德-茨纳耶克过程，第9章和图11.4（a））。[5]他们还通过弯曲时空的计算发现，在提取能量时，电流从黑洞极点附近流入视界（表现为正电荷粒子落进去），而从赤道附近流出（表现为负电荷粒子落进去）。黑洞仿佛是一个电路的一部分。

计算还说明，黑洞似乎也是电路中的电压发生器（图11.4（b））。这台电压发生器驱使电流从视界赤道流出，又将磁力线驱赶到远离黑洞的地方，然后将等离子体（导电热气体）驱赶到黑洞旋转轴附近的其他电力线上，最后又将这些电力线赶下来进入视

界。这些磁力线是电路的导线，等离子体是从电路提取能量的负载，能源则是旋转的黑洞。

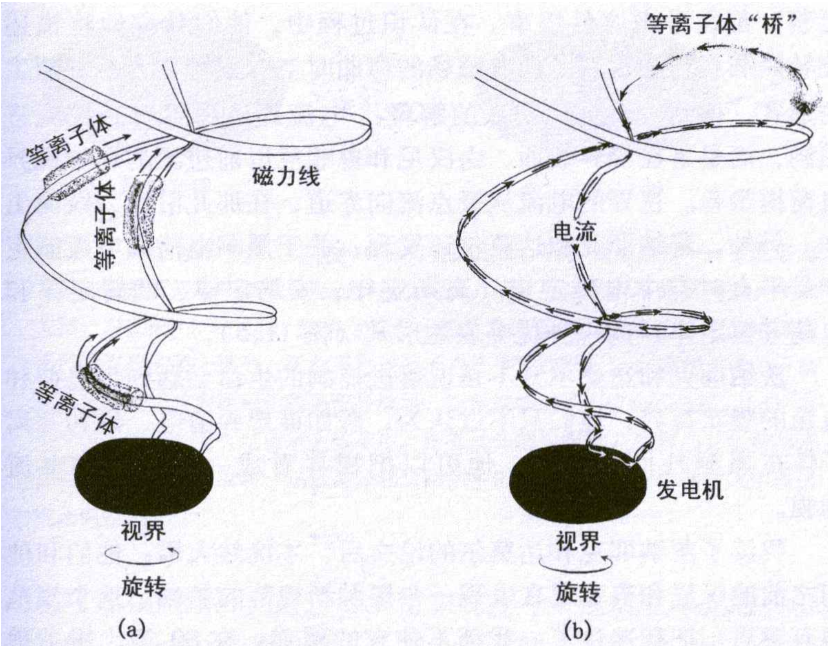



图11.4　旋转磁化黑洞赖以产生喷流的布兰福德-茨纳耶克过程的两种观点。（a）黑洞的旋转产生空间旋涡，迫使穿过黑洞的磁场旋转，旋转磁场的离心力将等离子体加速到很高的速度（与图9.7（d）对比）。（b）磁场和空间旋涡一起在黑洞极点和赤道间产生巨大电压，结果黑洞成为电压发生器和发动机。电压驱动回路产生电流，回路将黑洞能量带给等离子体并将它们加速到很高的速度。

照这点看（图11.4（b）），使等离子体加速并形成喷流的动力是由电路带来的。根据第9章的观点（图11.4（a）），动力则来自飘来飘去的旋转磁力线。两个观点不过是对同一件事情的不同考虑方式。在两种情况下，动力最终来源都是黑洞的旋转。人们可以凭自己的兴趣认为动力来自电路还是来自旋转的磁力线。

电路的观点，虽然以标准的弯曲时空的物理学定律为基础，却完全是不曾料想过的，而通过黑洞的电流——从极点附近流入，从赤道附近流出——似乎也太奇怪了。1977年和1978年间，茨纳耶克和达

莫尔（也是一个研究生，但不在剑桥，而在巴黎）都在考虑这件怪事。在认识过程中，他们独立地将描述旋转黑洞和它的等离子体和磁场的弯曲时空方程转化为一种陌生的形式，得到一个生动诱人的解释：[6]电流到达视界时并没流进黑洞，而是落在视界表面，由汉尼和鲁菲尼以前想象的那种视界电荷携带着。视界的电流从极点流向赤道，在那儿沿磁力线流出来。另外，茨纳耶克和达莫尔还发现，关于黑洞电荷和电流的定律是平直时空中电磁定律（高斯定律、安培定律、欧姆定律和电荷守恒定律）的一种优美表达形式（图11.5）。

茨纳耶克和达莫尔并不是说落进黑洞的生命会遇到带电荷和电流的膜状视界；他们只不过认为，假如谁想弄清电、磁和等离子体在黑洞外面的行为，他可以把视界看成一张带电和电流的膜。

我读了茨纳耶克和达莫尔的论文后，才恍然大悟：他们和他们之前的汉尼和鲁菲尼在发现一个黑洞新规范的基础。这个规范很有意思，把我迷住了。我挡不住它的诱惑，在80年代用了很多时间与普赖斯、麦克唐纳（Douglas Macdonald），雷德蒙特（Ian Redmount）、孙为默〔音〕、克罗里（Ronald Crowley）等人把它修饰了一下，写成一本书：《黑洞：膜规范》。[7]

写进膜规范的黑洞物理学定律完全等价于对应的弯曲时空规范的定律——只要我们将注意力局限在黑洞的外面。从而，对一切可能在黑洞外面进行的实验和观测——包括地球上的一切天文观测，这两个规范的预言是完全相同的。我发现，同时把握两个规范（膜的和弯曲时空的），在两者之间经历埃舍尔式的思路转移，对思考天文学和天体物理学问题是很有好处的。当我星期天考虑黑洞脉动时，弯曲时空规范可能更适用，因为它的视界是由弯曲虚空的时空构成的。当我星期一考虑黑洞喷流时，膜规范也许更适用，因为它的视界是一张带电的膜。由于两个规范的预言保证是一样的，不管星期几，哪个规范能适合我的需要，我就可以用哪个。[5]

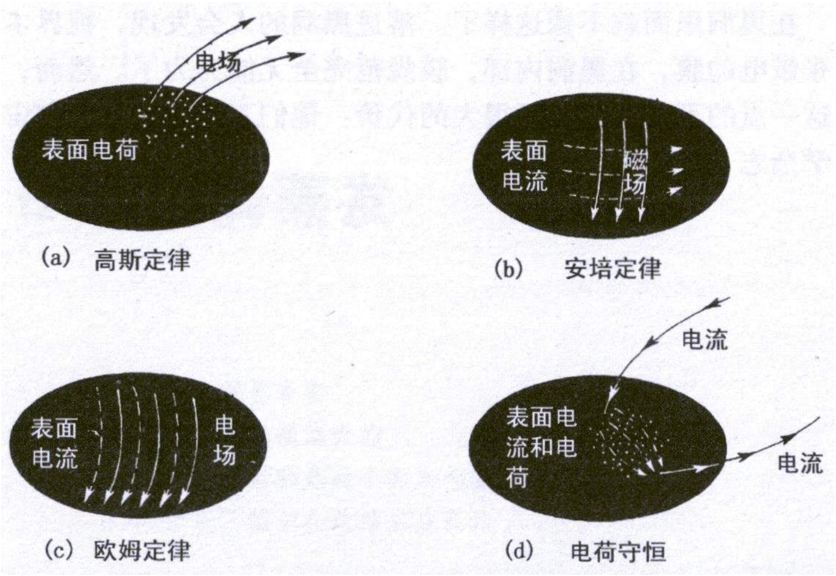



图11.5　黑洞膜状视界上电荷和电流的有关定律：（a）高斯定律——视界表面电荷的数量正好能终结所有与视界相交的电力线，所以它们不能延伸进入黑洞内部；对比图11.3。（b）安培定律——视界表面电流的总量正好能终结磁场平行于视界的那一部分；因此视界以下不存在平行磁场。（c）欧姆定律——表面电流正比于与表面相切的那部分电场，比例常数是377欧姆的电阻。（d）电荷守恒定律——没有电荷消失或产生。所有从外部宇宙进入视界的正电荷都落在视界上，在表面流动，然后离开它又回到宇宙（表现为负电荷落下来中和这些正电荷）。

在黑洞里面就不像这样了。落进黑洞的人会发现，视界不是一张带电的膜，在黑洞内部，膜规范完全无能为力了。然而，发现这一点的下落者付出了很大的代价：他们不可能在外面宇宙的科学杂志上公布他们的发现。

[1]　库恩的“规范”（paradigm），在一些哲学译著里译为“范式”。（汉译本《科学革命的结构》，李宝恒、纪树立译，上海科学技术出版社，1980。）——译者

[2]　埃舍尔（Maurits Cornelis Escher，1898～1972）喜欢以图画表现“不可能存在的世界”，从这点说，他大概是世界上最特殊的画家。他的画都充满了浓厚的数学趣味，用他的话说，“是为了传达思维的一条特殊线索……最终使艺术步入数学领域。”荷兰数学家恩斯特（Bruno Ernst）为他的这位同胞写了一本有名的《M.C.埃舍尔的魔镜》，带我们用数学眼光去欣赏他那些神奇的图画（这里看到的是他1961年的石版画《瀑布》）。——译者

[3]　比较第1章最后一节，“物理学定律的本质”。

[4]　“极化”在这里的意思与“极化引力波”和“极化光子”（第10章）有所不同。

[5]　“星期天”、“星期一”的说法，在西方是有传统的。英国诗人布拉斯韦特（Richard Brathwait，1588？-1673）在拉丁文诗《巴拿马日记》（1638）中写道：“我看见一个清教徒，星期一吊起他的猫，等着星期天逮老鼠。”更有趣而且与物理学关系更密切的是布拉格（W.L.Bragg）对玻尔的量子化概念的笑话：“在这个理论中我们似乎该在每周一、三、五用经典定律，在二、四、六用量子定律。”M.Bartusiak还有一本天文学科普书名叫《星期四的宇宙》（1986）。——译者

# 第12章　黑洞蒸发

视界裹在

慢慢消失的

辐射和热粒子的大气里，

黑洞在收缩然后爆炸

1970年11月的一个晚上，霍金正准备睡觉，忽然有了一个想法，它来得那么急，令他差点儿喘不过气来，他还从没遇到过一个思想来得这么快的。[1]

睡觉对霍金来说也真不容易。他患了肌萎缩性脊髓侧索硬化（ALS），支配肌肉的神经逐渐被破坏，一块块肌肉失去了活力。他两腿颤栗着慢慢地移动，刷牙时还得用一只手撑着桌台；他紧紧抓着床柱，脱去衣服，然后艰难地套进睡衣，爬上床。那天晚上，他比平常动作还慢，因为满脑子都是那个思想。这思想令他狂喜，但他没告诉妻子简，那会挨骂的，因为她满以为他会专心去睡觉。

那一夜，他醒着躺了好几个小时，睡不着。他的思维还徜徉在那个思想的枝枝叶叶上，还在寻找它与其他事物的联系。

这个思想是一个简单问题引发的。当两个黑洞碰撞结合成一个黑洞时，会产生多少引力辐射（时空曲率波）？霍金已经大概知道，最后那个黑洞从某种意义说比原来两个黑洞之“和”更大，但那是什么意义呢？关于产生了多少引力辐射，它又告诉他什么呢？

于是，在准备睡觉时，他想到了。突然，一系列的图景在他头脑里合成，产生了那个思想：更大的是黑洞视界的面积。他确信这一点，景象和图画已经形成了一个不容置疑的数学证明。不论原来两个黑洞质量多大（相同或大不相同），不论黑洞如何旋转（同向、反向或是根本不转动），也不论它们如何碰撞（正碰还是斜碰），最终黑洞视界的面积一定总是大于原来黑洞视界面积之和。那又怎么样呢？霍金的头脑还在这个面积增加定理中徜徉时就已经认识到，那太了不起了。

首先，最后的黑洞为了有更大的视界面积，一定要有很大的质量（或等价地说，很大的能量），这意味着作为引力辐射喷射出去的能量不太多。但“不太多”也不是太少。霍金通过把他新的面积增加定理与用面积和自旋表达的描述黑洞质量的方程结合，计算出原来两个黑洞质量的50%可以转化为引力波能量，只为最后那个黑洞留下50%的质量。[1]

在那个11月的不眠之夜后的几个月里，霍金又发现了他那思想的另一些枝叶。最重要的也许是他为下面这个问题找到了一个新答案：当黑洞是“动态”的时候，也就是，当它大幅度振动时（在碰撞中这是一定会发生的），或者当它快速增长时（当它最初由坍缩恒星产生时，这也是可能的），该如何定义黑洞视界的概念？

准确而成功的定义是物理学研究的基础。闵可夫斯基只是在定义了两个事件的绝对间隔后（卡片2.1）才发现，虽然空间和时间是“相对的”，但可以统一为一个“绝对的”时空。爱因斯坦只是在定义了自由下落粒子的轨迹是直线后（卡片2.2），才发现时空是弯曲的（图2.5），从而才创立了他的广义相对论。霍金也是在定义了动态黑洞的视界概念后，才能和别人去探索当黑洞受碰撞或下落碎屑的打击时，它会如何改变？

1970年11月以前，大多数物理学家都跟着彭罗斯，[2]认为黑洞视

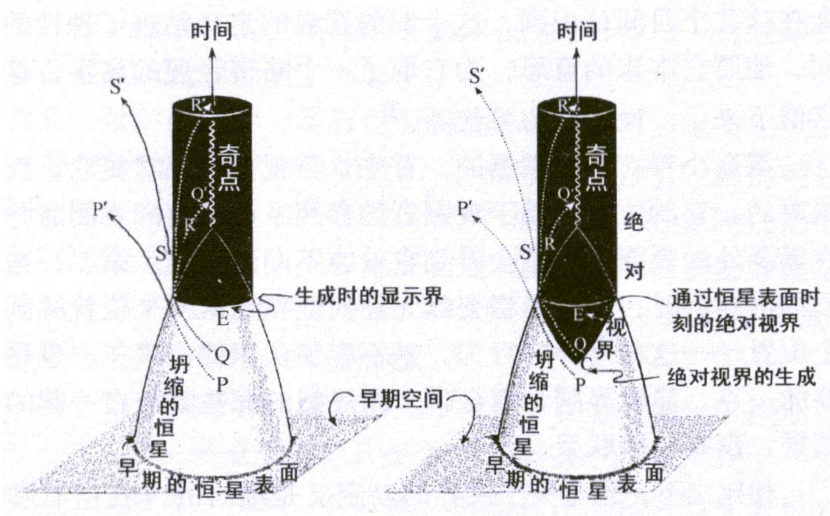
界是“试图逃逸黑洞的光子最后被引力拉下来的地方”。霍金在这几个月间认识到，这个旧的视界的定义钻进了理性的死胡同。他照它本来的意思，为它取了一个略带轻蔑的名字，这个名字留下来了。他称它为显视界。[2]

霍金小看它是有根据的。首先，显视界是相对概念，而不是绝对的。它的位置依赖于观测者的参照系；下落的观测者与静止在黑洞外的观测者可能会看到它处在不同的位置。第二，当有物质落进黑洞时，显视界将突然无任何征兆地从一个位置跳到另一个位置——这种奇异的行为，是不容易认识的。第三，也是最重要的一点，显视界同为霍金带来新思想的那些凝结在一起的智力图景，没有任何联系。

相比之下，霍金关于视界的新定义是绝对的（在所有参照系中都相同），不是相对的，所以他称它为绝对视界。霍金认为，绝对视界很优美。它有一个优美的定义：它是“时空中能否向遥远宇宙发送信号的事件之间的分界（视界外的事件能发送，而视界内的事件不能）。”[3]它还有一个优美的演化：当黑洞吞噬物质或与另一黑洞或其他事物碰撞时，绝对视界将以一种光滑、连续而不是突然、跳跃的方式发生形状和大小的改变（卡片12.1）。

卡片12.1　新生黑洞的绝对视界和显视界

下面的时空图描绘了球状恒星形成球状黑洞的坍缩，请与图6.7比较。点线是外出的光线，换句话说，它们是光子的世界线（通过时空的轨迹）——这种最快的信号可以径向向外发送到遥远的宇宙。对于“理想的逃逸”，我们理想化地认为光子不被任何恒星物质吸收和散射。




显视界（左图）是想逃脱黑洞的外出光线（如向外的QQ′和RR′）被拉向奇点的最外边界。显视界是恒星表面收缩经过临界周长时在E处突然完全生成的。绝对视界（右图）是能向遥远宇宙发送信号的事件（如事件P和S，沿光线PP′和SS′发送信号）和不能向遥远宇宙发送信号的事件（如Q和R）之间的分界。绝对视界在事件C的恒星中心生成，比恒星收缩到临界周长早一些。绝对视界生成时只是一点，然后像吹气的气球那样逐渐膨胀。当恒星收缩到临界周长（圆E）时，它也完全出现在恒星表面，这时不再扩张，以后就与突然形成的显视界一致。

更重要的是，绝对视界完全符合霍金的新思想：

霍金从凝结在他头脑中的图景看到，绝对视界（而不一定是显视界）的面积不仅在黑洞碰撞和结合时增大，而且在黑洞诞生时，在物质或引力波落下来时，在宇宙的其他事物的引力掀起潮汐时，在从它外面的空间旋涡中提取旋转能量时，都会增大。实际上，绝对视界的面积几乎总是增大，而永远不会减小。物理原因很简单：黑洞遭遇的任何事物都穿过它的绝对视界向内发送能量，任何能量都无法回到外

面来。由于所有形式的能量都产生引力，这意味着黑洞引力在不断加强，因而相应地它的表面积也不断地增加。

更准确地说，霍金的结论是：（在任何人的参照系中），在任何空间区域和时刻测量所有黑洞的绝对视界的面积，并把这些面积加到一起得一个总面积。然后，你可以等任意长的时间再测量这些绝对视界的面积并把它们加起来，假如在两次测量间没有黑洞从这一空间区域的“围墙”转移出去，那么视界的总面积不会减少，而几乎总会增加，至少增加一点儿。

霍金很清楚地知道，不论选择哪种视界的定义，是绝对的，或者显的，都不会以任何方式影响对人类或其他生物可能进行的任何实验结果的预言。例如，它不会影响对在黑洞碰撞中产生的引力波（第10章）的预言，也不会影响对落进黑洞视界的热气体发出的X射线数量（第8章）的预言。但是，定义的选择却关乎理论物理学家从爱因斯坦广义相对论方程演绎黑洞行为特征是费力还是轻松。在理论家用以指导研究的规范里，他所选择的定义将成为决定性的工具。它影响他们的思维图景，影响他们在与别人交流时说的话，也影响他们直觉的飞跃。在这一点上，霍金相信，新的绝对视界因它连续增长的面积，比旧的不连续跳跃的显视界更优越。

思考绝对视界并发现它们面积增加的物理学家，史蒂芬·霍金不是第一个。牛津大学的彭罗斯和加拿大艾伯塔大学的伊斯雷尔在霍金那个11月的不眠之夜以前就已经做过了。[4]霍金的发现实际也在很大程度上靠了彭罗斯打下的基础（第13章）。然而，不论彭罗斯还是伊斯雷尔，都没认识到面积增加定理的意义和力量，他们也没发表这个结果。为什么呢？他们的思想死抱着显视界作为黑洞表面，而把绝对视界当做某个无关紧要的辅助性概念，从而也不认为绝对视界面积的增加有多大意思。跟着我们这一章，你会看到他们犯了一个多么可怕的错误。

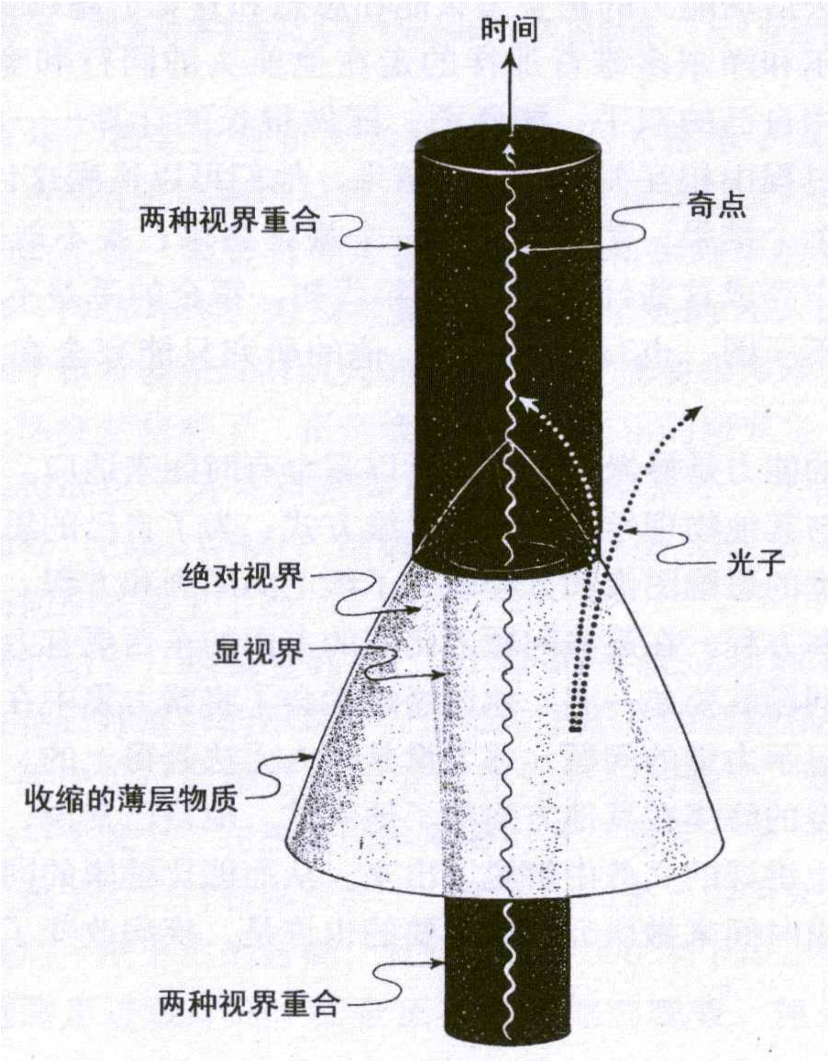
为什么彭罗斯和伊斯雷尔那么喜欢显视界呢？因为它曾在一个惊人发现里充当主角。那是彭罗斯1964年的一个发现：广义相对论定律迫使每个黑洞在中心有一个奇点。[5]我将在下一章讨论彭罗斯的发现和奇点的本质。我现在主要说的是，显视界显示了威力，而彭罗斯和伊斯雷尔被这威力蒙蔽了，没能想到要放弃它作为黑洞表面的定义。

他们特别不能想象放弃显视界而赞同绝对视界。为什么呢？因为绝对视界似乎自相矛盾地违背了我们信奉的结果不得先于原因的观念。当物质向黑洞落下时，绝对视界就开始增长（“结果”），而这时物体还没有到达它（“原因”）呢。视界在期待中增长，物质马上会被吞没，黑洞引力也将随之而增强（卡片12.2）。

彭罗斯和伊斯雷尔知道这个表面的矛盾是从哪儿来的。正是那个绝对视界的定义依赖于未来发生的事情：信号最终能否逃向遥远的宇宙。用哲学名词来说，这是一个目的论的定义（依赖于“最终原因”的定义），它使视界演化也是目的论的。由于现代物理学中极少出现目的论观点，所以彭罗斯和伊斯雷尔会怀疑绝对视界的价值。

卡片12.2　吸积黑洞显视界和绝对视界的演化[6]

下面的时空图说明了显视界的跳跃性演化和绝对视界的目的性演化。在某一初始时刻（近图底的一张水平面上），一个非旋转老黑洞为薄层球状物质外壳所包围。外壳像橡皮气球，而黑洞像气球中心的一个陷坑。黑洞引力作用在外壳（气球），使它收缩并最终将它吞没（就像气球落进陷坑）。显视界（向外的光线——图中点线——的最后一道屏障）在收缩外壳到达最后黑洞临界周长位置的瞬间突然发生不连续的跳跃，绝对视界（能否向遥远宇宙发送光线的事件的分界）在黑洞吞没外壳前开始扩张，在扩张中等待吞噬，然后，当黑洞吞没外壳时，它也停在跳跃的显视界的位置。




霍金是大胆的思想家，假如他感觉那些激进的新方向是对的，他会比大多数物理学家更乐意走上那些方向。对他来说，绝对视界的"滋味"不错，尽管有点儿"烈"，他还是喜欢，而且有了回报。在几个月里，他和哈特尔根据爱因斯坦广义相对论定律导出了一系列美妙的方程，描绘在吞没下落的物质碎片和引力波之前，在受其他物体引力作用之前，绝对视界是如何连续而光滑地扩张和改变形状的。[7]

1970年11月，霍金作为物理学家才刚迈出满意的一步，他已经有过一些重大发现，但还没成为主角。随着这一章的脚步，我们会看着他成为一名主角。

失去活动能力的霍金怎么能在思想和直觉上超越像彭罗斯、伊斯雷尔和泽尔多维奇那样的走在他前头的同行和竞争者呢？他们能用自己的双手，能画图，能做很长的计算——那些计算记录着过程中相互关联的一些结果，他们可以追溯这计算过程，检查一个个结果，然后综合成一个最终结果；我不敢想象谁能在头脑中完成这些计算。到70年代初，霍金的手差不多已经废了，画不了图，也写不了方程。他的研究只能完全在头脑里进行。

手的能力是慢慢丧失的，所以霍金有时间来适应。他逐渐练就一种与其他物理学家不同的思维方式：为了自己的思考，他以新的直觉的思维图像和方程取代了纸上的图画和方程。霍金的思维图景和方程，在某些问题上比旧的书面的东西更有力量，但对另外的问题就要差一些。他还慢慢学会了将精力集中在他的新方式更能显示力量的问题，那力量是别人无法赶得上的。

霍金的瘫痪在其他方面帮了他的忙。他自己常说，它让他从为大学生讲课的义务中解脱了出来，从而能比健康的同事们有更多的自由时间来做研究。更重要的也许是，疾病改变了他的生活态度。



1980年，霍金与妻子简和儿子Timonthy在英国剑桥。[K·索恩摄]

1963年，霍金上剑桥大学研究生院不久就患了ALS。ALS是一类运动神经元疾病的总称，大多数患者很快就会死去。想到只有几年的生命，霍金首先失去了对生活和物理学的热情。然而，1964～1965年冬，发现他患的是一种罕见的ALS病例，它逐渐破坏中枢神经系统对肌肉的支配能力，需要很多年，而不是几年。生活突然精彩了，霍金像一个健康快乐的研究生一样，带着从未有过的巨大活力和热情又回到物理学来了。生活重新开始，他和简（Jane Wilde）结婚了。简是他患ALS后不久认识的，得病初期就爱上了她。

与简的结合，是霍金在60、70和80年代成功和幸福的基础。她在他身体遭遇的不幸中为他带来了正常的家庭和生活。

我一生中见过的最幸福的笑是在史蒂芬的脸上。那是1972年8月的一天晚上，在法国阿尔卑斯山下。那天，简、我和他们的两个大孩子罗伯特和露茜游了一天下山回来。因为太笨，我们错过了最后一趟下山的雪橇，只好步行1 000米下山。当简、罗伯特和露茜走进饭厅时，霍金正在摆弄他的晚餐。他先还在替我们着急，看到他们进来时，却忍不住大笑起来，眼泪都流出来了，饭也吃不下了。

霍金的手脚都动不了，然后又逐渐失去了声音。1965年6月，我们第一次见面时，他拄着根手杖走路，声音只是略有颤抖。1970年，他得靠四腿的架子才能走路。到1972年，他只能坐在自动轮椅上，而且基本上不能写字了，但还能较轻松地自己进食，大多数地道说英语的人还能听懂他说话，当然有点儿困难。1975年，他不能自己进食了，也只有习惯了他讲话的人才能听懂他说什么。1981年，除非在绝对安静的屋子里，不然，他的话，我听起来也很费劲；只有长期同他在一起的人才会觉得容易些。到1985年，他的肺不能自动排气，需要切开气管，通过有规则的吸气清除气流障碍。手术的代价太高了：他完全失去了声音。他只得靠为他设计的一台计算机语音合成器来说话，抱歉的是说话迟钝，而且带着美国口音。他通过握在手上的简单开关控制计算机，在屏幕上打出一串串单词，然后用开关选出他需要的词组成句子。这是一个痛苦缓慢的过程，却很有效。他一分钟最多能造一个简单句子，但他的句子从合成器读来还是清楚的，而且很优美。

说话能力退化了，霍金学会了把每个句子都存起来。他找到了一种比他患病初期更清楚、更简洁的思想表达方式。随着清楚而简洁的表达，他的思想也更清晰，对同事们的影响也更大——但似乎也越来越费解了：有时，当他提出对某个深刻问题的判断时，我们这些同行要在想很久、做许多计算之后才能确定他是在猜想，还是已经有了强有力的证据。有时候他不告诉我们，而我们偶尔也怀疑他是不是在拿他绝对独特的思想跟我们开玩笑。毕竟，他还保留着在牛津读大学时那种讨人喜欢的顽皮和即使在患难时也没离开过他的幽默。（在支气管手术前，我已经开始难得听懂他的话了，有时得反反复复对他说：“史蒂芬，我还是没听懂；请再说一遍。”他有点儿泄气，但还是不断重复，直到我恍然大悟：原来他在给我讲一个精彩的异乎寻常的小笑话，当我终于笑了，他也愉快地笑了。）

## 熵

我在上面称赞了霍金在思想和直觉上超越同行竞争者的能力，但我现在应该承认，他并不总是在赢，也有输的时候。他最大的一次失败可能是败在惠勒的研究生贝肯斯坦（Jacob Beken-stein）手下。但我们将看到，在那次失败中，霍金获得了一个更大的胜利：发现黑洞蒸发。本章剩下的篇幅，就是讲这一发现的曲折经过。

霍金失败的战场在黑洞热力学。热力学是一组关于大量原子的随机的统计行为的定律，如组成房间里的空气的原子或组成整个太阳的原子。原子的众多统计行为中，包括由热引起的随机跳跃，相应地，热力学定律也包括关于热的定律，因此才有热力学这个名称。

在霍金发现面积定理的前一年，普林斯顿惠勒小组里的19岁研究生克里斯托多罗（Demetrios Christodoulou）注意到，描述黑洞性质缓慢变化（如它缓慢吸积气体）的方程很像某些热力学方程。[8]它们之间的相似是很明显的，但除了认为巧合外，找不到更多的理由。

霍金的面积定理又加强了这种相似：面积定理很像热力学第二定律。实际上，在本章前面的表述中，只要把“视界面积”换成“熵”，面积定理就变成了热力学第二定律：（在任何人的参照系中），在任何空间区域和时刻测量区域内所有事物的总熵。然后，你可以等任意长的时间再来测量，假如在两次测量间没有事物从你的空间区域的“围墙”跑出去，那么总熵不会减少，而几乎总会增加，至少增加一点儿。

那个增加的“熵”说的是什么东西呢？它是一定空间区域的“随机性”的总量。熵增加意味着事物在不断地变得越来越随机。

更准确些说（卡片12.3），熵是一定空间区域内所有原子和分子在不改变区域宏观表现情况下的分布方式的数目的对数。[4]如果原子和分子可能的分布方式多，微观的随机性就大，熵也就大。

熵增加定律（热力学第二定律）力量很大。举例说，假定一间屋子里有空气和几张皱巴巴的报纸，它们包含的熵小于报纸在空气中燃烧形成二氧化碳、水蒸气和一点儿灰后所包含的熵。换句话说，当屋子里原来是空气和报纸时，分子和原子的随机分布方式比最后在空气、二氧化碳、水蒸气和灰的情况下少。这也是为什么纸很容易点着而自然燃烧，而燃烧却不容易自然地倒过来从二氧化碳、水、灰和空气还原成纸。熵在燃烧中增加，在还原中减小，所以燃烧会发生，而还原却不能。

1970年11月，霍金立即就注意到了热力学第二定律和他的面积增加定理之间的相似性，但他显然认为这只是一种巧合。他想，谁要是说黑洞视界就是某种意义的黑洞的熵，那他一定是疯了，至少是昏了头。是的，毕竟黑洞没什么随机的东西。黑洞倒是随机的对头，是简单性的化身。一旦黑洞处于一种宁静状态（通过发出引力波，图7.4），它就完全“无毛”了：一切性质都由三个数决定：质量、角动量和电荷。黑洞无论如何没有随机性。

卡片12.3　玩具屋的熵

一间正方形的玩具屋里有20个玩具。屋子的地板铺着100块大瓷砖（每边10块）。爸爸打扫完屋子，把玩具放在最北的那行地砖上。他不在乎哪个玩具放在哪块砖，所以玩具完全是随机堆在一起的。随机性的一种度量是它们有多少种堆放方式（不论哪种方式，爸爸都一样满意），也就是，20个玩具放在北边那行的10块地砖上所能有的分布方式的数目，它是$10\times10\times\ldots\times10$，也就是$10^{20}$，因为每个玩具都有10种方式。

这个$10^{20}$就是对玩具的随机性的一种描述。然而这是一个很难把握的描述，因为$10^{20}$太大了。更容易把握的是它的对数，也就是多少个10的因子乘起来能得$10^{20}$，那就是20。玩具在地砖上堆放方式的数目的对数，就是玩具的熵。

这时候，孩子进屋来玩儿，把玩具扔得到处都是，然后他又走了。爸爸回来看见一团糟。现在的玩具比先前更混乱了，它们的熵增加了。爸爸不管哪个玩具在哪儿，他看到的是玩具随意地分散在整个屋子里。它们有多少种不同的分布方式呢？20个玩具分散在100块地砖上，有多少种方法？$100\times100\times\ldots\times100$，每个玩具100种，那么总数就是$100^{20}=10^{40}$，它的对数是40，于是孩子将玩具的熵从20增加到了40。

“那有什么，他爸爸接着会整理房间的，于是玩具的熵又减到20，”你大概会说，“这不就违反了热力学第二定律吗？”根本没有。爸爸可以通过整理将玩具熵减回来，但他的身体和屋子里的空气的熵却增加了：为把玩具放回原来的地方，爸爸得“燃烧”一些体内脂肪来获得能量。燃烧将有机的脂肪转化为无机的废物，如他在屋子里随机呼出的二氧化碳。结果，爸爸和屋子的熵的增加（原子和分子的可能分布数的增加）远远抵消了玩具熵的减小。

贝肯斯坦不服，[9]在他看来，黑洞的面积在某种深层意义上就是它的熵——或者更准确些说，是它的乘了某个常数的熵。贝肯斯坦论证说，假如不是这样，假如黑洞像霍金说的那样没有熵（没有任何随机性），那么黑洞就可用来减少宇宙的熵，这样就违背了热力学第二定律。我们只需要将从某个空间来的所有空气分子装进一个小口袋然后扔进黑洞就行了。口袋落进黑洞时，这些气体分子和它们携带的熵便从宇宙中消失了；假如黑洞不增加熵来补偿这些损失，那么宇宙的熵就减少了。贝肯斯坦认为，这样违背热力学第二定律是很不令人满意的。为了保住第二定律，黑洞必须拥有熵，在气体落进视界时它会增大；而在贝肯斯坦眼里，最有希望成为熵的候选者就是黑洞表面的面积。

根本不是那样的，霍金答话了。我们能通过把气体分子扔进黑洞而失去它们，当然也可以失去熵。霍金认为，这正是黑洞的本质，我们只能接受违反热力学第二定律的事实，那是黑洞性质要求的——除

此而外，它也没有任何严重的后果。例如，在通常情形，违反热力学第二定律可能允许制造永动机，但即使黑洞破坏了第二定律，永动机也是不可能的。这种破坏只是物理学定律的一个小小特例，有这些特例，物理学定律还是可以很好地存在下去。

贝肯斯坦还是不服。

全世界所有的黑洞专家都站在霍金一边——只有一个例外，那就是贝肯斯坦的导师，约翰·惠勒。他告诉贝肯斯坦，“你的思想够疯狂了，它可能是对的。”在导师的鼓励下，贝肯斯坦奋勇向前，加强了他的猜想。他估算了为保留热力学第二定律，在气体包落入黑洞时，黑洞的熵应正好增加多少；他还估算了，落进来的气体能增大多少黑洞的面积。根据这些粗略估计，他导出了熵和面积之间的一个关系，他认为这个关系可能总会满足热力学第二定律；熵近似地等于视界面积除以一个与量子引力定律（那时还没有呢）相关的著名面积，普朗克-惠勒面积，$2.61\times10^{-66}$平方厘米。[5]（在以下两章我们将学习普朗克-惠勒面积的意义。）对10个太阳质量的黑洞来说，熵是黑洞面积，11 000平方公里，除以普朗克-惠勒面积，$2.61\times10^{-66}$平方厘米，结果大概是$10^{79}$。

这是一个巨大的熵，代表着大量的随机性。这些随机性在哪儿呢？贝肯斯坦猜测在黑洞里。黑洞内部一定包含着大量的原子、分子或别的东西，所有这些东西都随机分布，可能分布方式的总数一定是$10^{10^{79}}$。[6]

废话！大多数前沿黑洞物理学家，也包括霍金和我都这样反应。黑洞内包含着一个奇点，没有原子，也没有分子。

然而，不管怎么说，热力学定律和黑洞性质之间的相似总是令人惊讶的。

1972年8月，黑洞研究的黄金年代正活跃的时候，全世界的主要黑洞专家和约50名学生相聚在法国阿尔卑斯山上，紧张地做一个月的讲习和联合研究。地方还是9年前（1963）我学广义相对论的那个莱苏什暑期学校，还是面对勃朗峰那绿油油的山坡（第10章）。[10]1963年我还是学生，现在，1972年，人家说我是专家了。早上，我们这些“专家们”互相交流，也向学生们讲过去5年的发现和我们现在努力的新方向。大多数下午的时间我们都在不断讨论新问题：诺维科夫和我关在小木屋里，想发现吸积到黑洞的气体发射X射线的规律（第8章）；而在学校休息室的长椅上，我的学生普雷斯和特奥科尔斯基在探讨旋转黑洞对小干扰是否稳定（第7章）；在我们上面50米的山坡上，巴丁、卡特尔和霍金在全神贯注地用爱因斯坦的广义相对论方程推导一组完整的黑洞演化定律。那真是难忘的田园诗，醉人的物理学。

月底，巴丁、卡特尔和霍金对黑洞力学定律的认识更牢了，这些定律与热力学定律有着惊人的相似。[11]实际上，只要以“熵”替代“视界面积”，以“温度”替代“视界表面引力”，我们就会发现，每一个黑洞定律都等同于一个热力学定律。[7]（所谓表面引力，粗略地说就是静止在视界上的人所感受的引力作用强度。）当贝肯斯坦（他是讲习班的50名学生之一）看到两组定律有那么完美的对应时，比以前更加相信视界面积就是黑洞的熵。相反，巴丁、卡特尔、霍金、我和其他专家却从这些对应中看出它严格证明了视界面积不可能充当黑洞的熵。假如是的话，表面引力就该充当黑洞的温度了，而温度不能是零。然而，热力学定律主张一切非零温度的事物都一定产生辐射（至少一点儿，家用取暖器就是这么工作的），而每个人都知道，黑洞不会发出任何东西。辐射会落进黑洞，但没有辐射能从黑洞跑出来。

假如贝肯斯坦能从他的直觉得到其逻辑结果，他会认定黑洞一定以某种方式具有有限的温度而且一定产生辐射，那么我们今天会把他看成了不起的先知。但贝肯斯坦不过是在瞎忙。他承认黑洞显然是不

能辐射的，但仍然顽固地相信他的黑洞熵。

[1]　霍金的面积增加定理允许任何黑洞质量都能以引力波形式发射出去，这似乎与我们的直觉矛盾。熟悉代数的读者可以从下面这个例子找到满意的答案：两个无旋转黑洞结合成一个更大的无旋转黑洞。无旋转黑洞的表面积正比于视界周长的平方，从而也就正比于黑洞质量的平方。这样，霍金的定理认为，初始黑洞质量的平方和一定大于*最后黑洞质量的平方。简单的代数计算可以说明，这个质量约束条件允许最后黑洞的质量小于原来两个黑洞质量之和，这样也就允许一定的初始质量作为引力波发射出去。（*显然是“小于”的笔误。即使$M^2 > M_1^2 + M_2^2$，仍然可能有$M < M_1 + M_2$，这两个条件可以确定最多能有多少质量转化成为引力波。——译者）

[2]　显视界更精确的定义见下面的卡片12.1。

[3]　在《时空的大尺度结构》里，霍金就称它是“事件视界”。——译者

[4]　量子力学定律保证了原子和分子的分布状态数总是有限而不会是无限的，物理学家在定义熵时常以它的对数乘以一个与我们无关的常数，$\log_e 10 \times k$，这里$\log_e 10$是10的“自然对数”，2.30258…，k是“玻尔兹曼常数”，$1.38062 \times 10^{-16}$尔格每摄氏度。我在全书都将忽略这个常数。

[5]　普朗克-惠勒面积公式为$G\hbar/c^3$，这里，$G = 6.670 \times 10^{-8}$达因·厘米$^2$/克$^2$是牛顿引力常数，$\hbar = 1.055 \times 10^{-27}$尔格·秒是普朗克量子力学常数，$c = 2.998 \times 10^{10}$厘米/秒是光速。相关问题见第13，14章的有关脚注和这些章节的讨论。

[6]　$10^{10^{79}}$的对数是$10^{79}$（贝肯斯坦猜想的熵）。注意，$10^{10^{79}}$是1后面跟$10^{79}$个0，就是说，0的数目与宇宙的原子一样多。

[7]　我们知道，热力学有四个定律，现在将这四个定律与相应的黑洞力学定律并列在下面：第零定律：系统平衡时，温度处处相同；视界在平衡状态下，表面引力处处相同。第一定律：能量守恒定律（当然都是满足的）。第二定律：熵永不减少；视界面积永不减少。第三定律：不可能通过有限步骤达到绝对零度（-273.15℃）；不可能通过有限步骤将黑洞表面引力减小到零。在不同场合，这些定律有不同的表述方式，但本质是相同的。——译者

# 黑洞辐射

事实上，黑洞能辐射的第一个线索来自泽尔多维奇，那是1971年6月，在莱苏什讲习班的14个月以前。然而，没人注意它。为此我感到惭愧，因为在他向根本性的新思想摸索着前进时，我是他的朋友，还跟他争论过。

那时，泽尔多维奇第二次请我到莫斯科在他的小组里做几个星期的研究。[12]两年前，我第一次去时，莫斯科住房正紧，泽尔多维奇为我在十月广场附近的夏伯罗夫卡街弄了一套宽敞的私宅。当我一些朋友正与他们的妻儿和父母挤在一间屋——是一间屋，而不是带一间卧室的套房——时，我却在独自享受着一间卧室，一间起居室，一间厨房，一台电视，还有精制的瓷器。而这一次我住的简陋多了，住在以前那所房子下一家苏联科学院旅店的单人房间里。

一天早晨6点半，我被泽尔多维奇的电话吵醒了。“快来我家，基普！我有了旋转黑洞的新想法！”我知道早点是吃不了了，咖啡、茶和馅饼（用牛肉末、鱼、白菜、果酱和鸡蛋做的一种酥油饼）都得等。我往脸上泼点儿凉水，套上衣服，抓起公文包，冲下五段楼梯来到街上，挤上一辆拥挤的无轨电车，然后又转车，在列宁墓的沃罗别夫斯科耶·绍瑟街2B号下了车，这是在克里姆林宫以南10公里的地方。紧邻的4号住着柯锡金（Alexei Kosygin），前苏联总理。[1]

我走进八尺高铁栅栏的一扇打开的门，来到一个四亩大的绿树荫荫的园子，周围绕着巨大的矮墩墩的房子，2B和连在一起的2A。墙上的黄色图画已经有些脱落。因为泽尔多维奇对前苏联核力量所做出的贡献（第6章），他得到2B的8间房，即二楼西南角的四分之一。在莫斯科，这套房子是很大的，有1500平方英尺。他跟夫人帕夫洛娃（Varara Pavlova）、一个女儿和女婿住在一起。

泽尔多维奇在门口等我，带着热情的笑容，后面房间里传出一家人忙碌的声音。我脱了鞋，套上门边堆着的拖鞋，跟他走进陈旧而舒适的起居室兼饭厅，屋里的沙发和椅子上堆满了东西；一面墙上挂着世界地图，泽尔多维奇去过的地方（伦敦、普林斯顿、北京、孟买、东京，还有很多）和前苏联政府因为害怕泄露核机密而不让他去的地方，都钉着彩色大头钉。

他两眼闪着光，[13]等我坐在屋子中央长长的饭桌旁，他宣布，“旋转的黑洞一定会辐射。离开的辐射将反作用在黑洞上，逐渐使旋转变慢，然后停下来。旋转没有了，辐射也就停止，而黑洞此后将永远处于无旋转的理想球形状态。”

我坚决地说：“这是我所听过的最疯狂的事情。”（我不喜欢和别人公开对抗，但泽尔多维奇却是在对抗中成熟起来的，他需要它，期待它；他把我带到莫斯科来，部分就是想让我充当他的争论对手，通过与我争论来检验他的思想。）我问，“你怎么会说出这么疯的话来？谁都知道，辐射能流进黑洞，但没有什么东西，也没有辐射能从那儿出来。”

泽尔多维奇解释了他的理由：“旋转的金属球会发出电磁辐射，所以同样的，旋转的黑洞应该发射引力波。”

我想，这是典型的泽尔多维奇式的证明。纯粹的物理直觉，没有任何基础，不过靠一些类比罢了。泽尔多维奇不太懂广义相对论，不会计算黑洞该做什么，所以他去计算旋转金属球的行为，然后断言黑洞会有类似行为，然后，大清早6点半把我叫来，检验他的论断。

不过，我已经见识过泽尔多维奇在比这更没有基础的情况下的发现了。例如，他1965年宣称，有山隆起的恒星在坍缩时会产生完全球状的黑洞（第7章），后来证明这个论断是对的，还预言了黑洞是无毛的。于是，我小心地问，“我不知道旋转的金属球会发出辐射。它

是怎么辐射的呢？”

“这种辐射很弱，”泽尔多维奇解释，“没人注意过，以前也没人预言。但是，它一定会发生。当电磁真空涨落刺激金属球时，它会辐射的。同样，当引力真空涨落擦过视界时，黑洞也会辐射。”

我在1971年真是太笨了，没能认识到这个论断的深刻意义，不过几年后它就清楚了。以前所有的黑洞理论研究都以爱因斯坦的广义相对论定律为基础，那些“黑洞不能辐射”的研究当然是不容置疑的。然而，我们理论家知道，广义相对论只是某个真正的引力定律的近似——它处理黑洞问题我们认为应该是很好的，但近似终归是近似。[2]我们确信，真正的定律一定是量子力学的，所以我们称它们是量子引力的定律。尽管这些量子引力的定律还只有些模糊认识，惠勒在50年代就得出了它们必然存在着引力真空涨落，一种小小的不可预料的时空曲率波动，即使时空完全没有物质，即使有人设法将引力波都从时空中拿走了，就是说，时空成了理想真空时，它仍然存在（卡片12.4）。泽尔多维奇说的，是他根据电磁学的类比预见了这些引力真空涨落会导致旋转黑洞辐射。“但是怎么辐射呢？”我还是很迷惑地问他。

卡片12.4　真空涨落

电磁波和引力波的真空涨落，相当于电子的“幽闭简并运动”。

回想一下（第4章），如果把电子限在一个小空间区域里，不论费多大气力让它慢慢停下来，量子力学定律都强迫它继续随机地不可预测地运动。这就是产生白矮星赖以抵抗引力挤压的压力的幽闭简并运动。

同样，如果有人想把电磁和引力振荡完全从某个空间区域拿走，他也是永远不会成功的。量子力学定律认为总还存在一些随机的不可预测的振荡，也就是随机的不可预测的电磁波和引力波。这就是（照

泽尔多维奇的观点）“刺激”旋转金属球和黑洞并导致辐射的真空涨落。

真空涨落不能通过拿走它们的能量而让它们停下来，因为总体说来它们本没有能量。在某些位置和某些时刻，它们可能有从别的地方“借来的”正能量，结果那些地方出现负能量。就像银行不会让债欠得太久，物理学定律迫使负能区域很快从正能的邻居那儿吸引能量，从而还原到零或得到些正能量。这样持续不断的随机的能量借还过程就驱动着真空涨落。

电子所限区域越小，它的简并运动越强（第4章）；同样，电磁波和引力波的真空涨落在小区域比在大区域更强，也就是，短波比长波的涨落强。在13章我们会看到，这对黑洞中心的奇点的性质有深刻意义。

在日常的物理中，电磁真空涨落是很普遍的，我们有良好的认识。例如，在荧光灯管的运行中，它们起着关键的作用。电荷激发灯管里的水银蒸气原子，然后随机的电磁真空涨落又刺激每个被激发的原子，使它在某一随机时刻以电磁波（光子）*形式释放出它的激发能量。由于最初认识这种发射时，物理学家还不知道它是真空涨落诱发的，所以称它是自发的发射。又比如，在激光器里，随机电磁真空涨落与相干的激光发生干涉（卡片10.3意义上的干涉），从而以一种不可预知的方式调节激光频率。这样，光子将在随机的不可预料的时刻从激光器发出来，而不像原来那样均匀地一个接着一个地出来——这种现象叫做光子闪动噪声。

与电磁波不同，引力的真空涨落还从来没有在实验上看到。凭90年代的技术和努力，应该能从黑洞碰撞中探测到高能的引力波（第10章），但不一定能发现微弱得多的真空涨落。

*涂在管壁的磷光吸收这个“初级”光子，然后发出“次级”光子，就是我们看到的光。

泽尔多维奇站起来，快步走到地图对面墙上1平方米的黑板前，一边画草图一边解释。他画的（图12.1）是一列流向一个旋转物体的波，从表面滑过，很快流走了。泽尔多维奇说，波可以是电磁的，旋转体是金属球；波也可以是引力的，旋转体为黑洞。

泽尔多维奇解释，流过来的波并不是“真正的”波，而是真空涨落。涨落的波从旋转体周围扫过时，就像一队溜冰者转弯儿，在外面的人转弯速度大，里面的人慢得多；同样，波的外面部分以很高的速度（光速）运动，而里面的部分比光慢得多，实际上比物体表面旋转还慢。[3]泽尔多维奇称，在这样的情况下，就像小孩儿快速甩动绳子为投石器加速那样，快速旋转的物体将抓着引力波，令它加速。加速过程中，波会从物体的旋转能中汲取一部分，将自己放大。长大的那部分波是带正能量的“真”波，而原来的没有长大的部分还是总能量为零的真空涨落（卡片12.4）。这样，旋转物体把真空涨落当成一种产生真实波动的催化剂，当成真实波动形态的模板。泽尔多维奇指出，真空涨落导致振动分子“自发地”发出光，也是类似的行为（卡片12.4）。

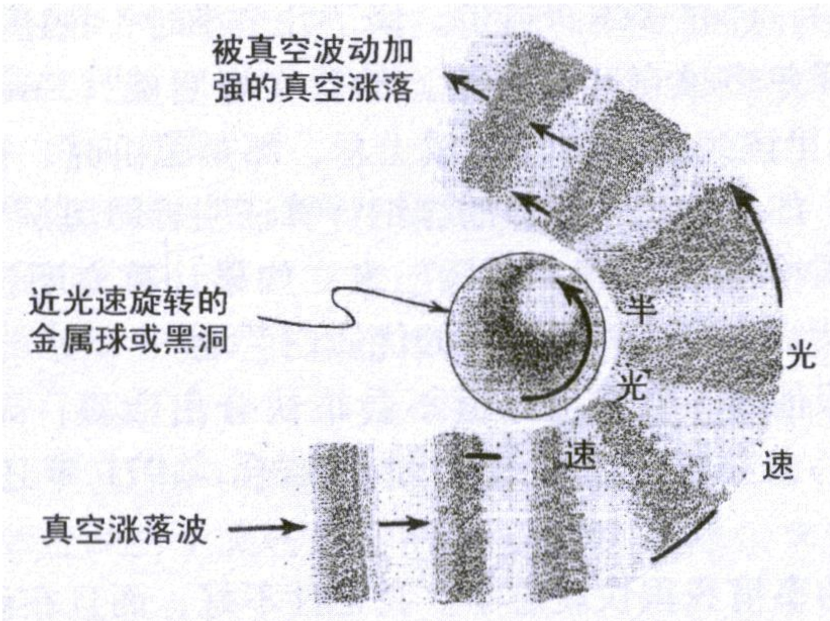



图12.1　真空涨落导致旋转物体辐射的泽尔多维奇机制。

泽尔多维奇告诉我，他已经用这种方法证明了旋转金属球会辐射，证明的基础是量子电动力学的定律——从量子力学与麦克斯韦电磁定律的结合产生出的一套成熟的定律。虽然他还没有同样地证明旋转黑洞的辐射，但通过类比他确信那是一定的。实际上他断言，旋转黑洞不但辐射引力波，也辐射电磁波（光子）、[4]中微子以及可能在自然界存在的所有其他形式的辐射。

我确信，泽尔多维奇错了。谈了几个小时，我还是不同意他的观点。泽尔多维奇和我打赌。他曾在海明威（Ernest　Heming-way）小说里读到过“白马”威士忌，那是很讲究、很地道的名牌威士忌。如果物理学定律的详细计算证明旋转黑洞要辐射，那么我从美国给他带一瓶“白马”来；如果计算表明没有这样的辐射，他就给我一瓶乔治时代的优质白兰地。

我愿跟他赌，但我也知道不会很快分出输赢，还得等更深刻地认

识广义相对论和量子力学的结合，1971年还没有人做得到。

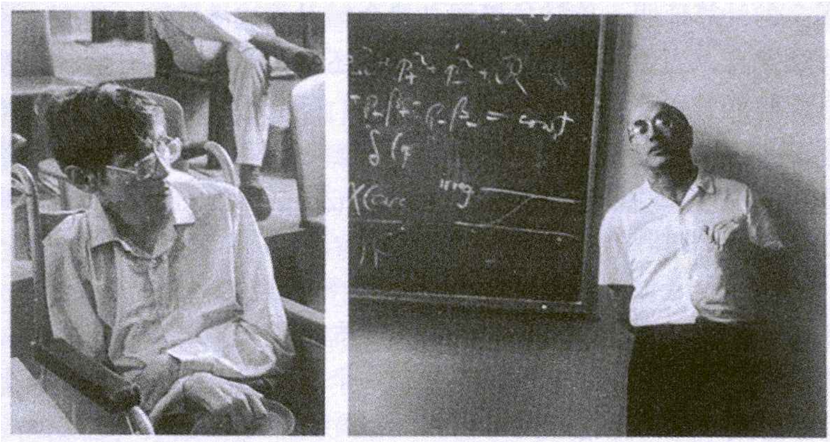
打赌的事情我很快就忘了。我记性不好，而且在研究别的东西。但泽尔多维奇没忘。在和我那次讨论的几星期后，他把论证写出来拿去发表。假如是别人写的，审稿人一定会拒绝的；他的论证试探性太强，不会让人接受的。但泽尔多维奇的名头太响，文章还是发了——几乎没人注意它。[14]黑洞辐射根本就是难以置信的。

一年后，在莱苏什讨论班上，我们“专家”依然没重视泽尔多维奇的思想。我想甚至没有谁提过一次。[5]

1973年9月，我又来到莫斯科，这次是陪霍金和他的夫人简。这是霍金上学以来第一次来莫斯科。他、简和泽尔多维奇（我们的前苏联主人）对霍金在莫斯科的特殊需要感到难办，就想到最好由既熟悉莫斯科，又是霍金和简的朋友的我来作伴，既做物理学会谈的翻译，又做导游。

我们住在罗西娅宾馆，离克里姆林宫附近的红场不远。虽然几乎每天都去各学校演讲或者参观博物馆，听歌剧，看芭蕾，我们还是在霍金的宾馆套房里，伴着窗外的圣巴西勒教堂，同许多主要的苏联物理学家进行了交流。他们一个个来到宾馆，向霍金表示敬意，与他交谈。

泽尔多维奇和他的研究生斯塔罗宾斯基（Alexi Starobinsky）是霍金宾馆的常客。霍金发觉他们也跟他一样吸引人。一次，斯塔罗宾斯基向他讲述了泽尔多维奇的旋转黑洞应该辐射的猜想，还讲了他和泽尔多维奇（以德维特、帕克等人以前的开拓性研究为基础）建立的量子力学与广义相对论的部分结合，然后说，这种部分结合证明了旋转黑洞的确会辐射。[15]泽尔多维奇很可能要赢我了。



左：1972年夏，S·霍金在莱苏什讲习班上听课。右：1971年夏，Y·B·泽尔多维奇在家中的黑板前。[K·索恩摄]

在莫斯科从别人的谈话了解的东西中，这一点是最令霍金感兴趣的。不过，他也怀疑泽尔多维奇和斯塔罗宾斯基结合广义相对论定律和量子力学定律的方式，所以，一回到剑桥，他就开始自己去结合，并用它来检验泽尔多维奇的旋转黑洞会辐射的论断。

同时，有几个美国物理学家也正做着同样的事情，其中有昂鲁什（William　Unruh，惠勒最近的学生）和帕奇（Don　Page，我的学生），他们在1974年分别以自己的方式初步证实了泽尔多维奇的预言：旋转的黑洞将发出辐射，直到所有旋转能耗尽，辐射才会停止。看来，我得承认自己输了。

[1]　绍瑟街从那时起改名为柯锡金街，建筑物也重新编了号。80年代末，戈尔巴乔夫曾住10号，在泽尔多维奇西边，隔着几道门。

[2]　见第1章最后一节，“物理学定律的本质”。

[3]　用专业术语说，外面部分在“辐射带”，而里面部分在“邻近带”。

[4]　回想一下，光子和电磁波是同一事物的不同方面；见卡片4.1中关于波粒二象性的讨论。

[5]　同时，米斯纳在美国证明，真实的波（与泽尔多维奇的真空涨落相反）能被旋转

黑洞以类似于图12.2的方式放大，这种放大过程——米斯纳称它为“超辐射”——激起了很大兴趣。照此看来，人们对泽尔多维奇的辐射缺乏兴趣就更令人深思了。

## 黑洞的收缩和爆炸

接着发生了一件爆炸性事件。[16]霍金先在英国的一个会上，然后又在《自然》杂志的一篇短文里宣布了一个惊人的预言，一个与泽尔多维奇、斯塔罗宾斯基、帕奇和昂鲁什矛盾的预言。霍金的计算证明，旋转的黑洞必然会辐射并减慢旋转。然而，计算还预言，当黑洞停止旋转时，辐射还不会停止。没有旋转，没有旋转能量，黑洞继续发出各类辐射（引力的、电磁的、中微子的），而在辐射时继续失去能量。我们知道，旋转的能量贮藏在视界外面空间的旋涡里，而现在失去的能量只能来自一个地方，那就是黑洞的内部。

同样令人惊奇的还有，霍金的计算预言，辐射谱（即每一波长的辐射总能量）完全像高温物体的热辐射谱。换句话讲，黑洞的行为仿佛在说它的视界具有一个有限的温度，而霍金算出这个温度正比于黑洞表面的引力。（假如霍金是对的）这确凿无疑地证明了巴丁-卡特尔-霍金的黑洞力学定律实际上就是另一形式的热力学定律，而且，正如贝肯斯坦两年前宣布的，黑洞具有正比于其表面积的熵。

霍金的计算还说明，一旦黑洞旋转慢下来，它的熵和视界的面积正比于质量的平方，而温度和表面引力正比于质量除以面积，也就是与质量成反比。因此，当黑洞持续辐射，将质量转化为外流能量时，它的质量下降，熵和面积下降，而温度和表面引力升高，黑洞收缩而变热，从结果看，它在蒸发。

一个刚从星体坍缩形成的黑洞（于是质量比2个太阳大）的温度很低，不超过绝对零度$3\times10^{-8}$度（0.03微开尔文）。因此，开始时蒸发很慢，要等$10^{67}$年（宇宙现在年龄的$10^{57}$倍）才会出现可以觉察的收缩。然而，随黑洞收缩和加热，它的辐射会越来越强，蒸发也将加快。最后，黑洞质量减小到几千到1亿吨之间的某个量（我们还不知

道确切的数量），视界收缩到原子核大小的若干分之一时，它将达到极高的温度（1万亿到10万万亿度），从而在几分之一秒内发生猛烈的爆炸。

在广义相对论与量子论的结合上，霍金与别人不同。世界上的十几位专家都确信霍金犯了错误。他的计算违反了那时所知的关于黑洞的一切事情。也许他的结合错了；也许他的结合对了，但计算错了。

接着的几年里，专家们常检验霍金和他们自己的结合形式，检验霍金和他们自己关于来自黑洞的波的计算。他们一个个逐渐走近霍金的结果；在这个过程中巩固了广义相对论与量子论的部分结合，形成了一组新的物理学定律，被称为弯曲时空的量子场的定律，因为在产生这些定律的结合中，黑洞被认为是非量子力学的、广义相对论的弯曲时空物体，而引力波、电磁波和其他类型的辐射被认为是量子场——也就是遵从量子力学的波，行为有时像波，有时像粒子（见卡片4.1）。（广义相对论与量子论的完全结合，即完全正确的量子引力定律将把包括黑洞的弯曲时空在内的一切事物都看成是量子力学的，也就是服从测不准原理（卡片10.2）、波粒二象性（卡片4.1）和真空涨落（卡片12.4）。下一章我们会看到这样的完全结合和它的某些应用。）

没有任何实验指导我们去选择，我们如何能够同意弯曲时空的量子场的基本定律呢？没有实验检验，专家们又如何能够几乎肯定地宣称霍金是对的呢？他们的根据是，量子场的定律与弯曲时空的定律应该完全一致地融合在一起。（假如融合不完全一致，那么在某种运用中，物理学定律可能做出一种预言，例如黑洞永不辐射；而在另一种运用中，它可能做出不同的预言，例如，黑洞必然总是在辐射。可怜的物理学家不知道该相信哪个，不知道该怎么做。）

新的综合的定律必须与没有量子场的弯曲时空的广义相对论定律和没有时空曲率的量子场定律一致。这一点与完美结合的要求，类似

于纵横字谜的行与列必须完全一致，[17]它们几乎完全决定了新定律的形式。[1]如果定律确实可以一致地融合起来（如果物理学家认识宇宙的方法是合理的，那是一定能做到的），它们也只能以新的和谐的弯曲时空的量子场定律所描述的方式进行。

物理学定律应一致融合的要求，通常是寻找新定律的工具。不过，这种一致性要求很少像在弯曲时空的量子场的场合下显示出那么大的威力。例如，在爱因斯坦建立他的广义相对论定律时（第2章），一致性的考虑不能也没有告诉他从哪儿起步，没有告诉他引力来自时空的曲率；这个出发点主要来自他的直觉。然而，有了这个基础，广义相对论定律在弱引力条件下与牛顿引力定律相容的要求和在无引力条件下与狭义相对论相容的要求，几乎惟一决定了新定律的形式；实际上，它是爱因斯坦发现场方程的关键。

1975年9月，我第五次访问莫斯科，为泽尔多维奇带了一瓶“白马”威士忌。现在，西方的专家们都同意霍金是对的，相信黑洞会蒸发，而我却惊讶地发现，莫斯科没人相信霍金的计算和结论。虽然在1974年和1975年间，发表了几个以新的完全不同的方法导出的对霍金论断的证明，但它们在前苏联并没有产生什么影响，为什么？因为泽尔多维奇和斯塔罗宾斯基这两位大专家不相信：他们还在认为，辐射黑洞失去旋转后，辐射也就停止了，从而不可能完全蒸发。我同泽尔多维奇和斯塔罗宾斯基争论过，但没有结果；他们对弯曲时空的量子场远比我懂得多，尽管（像通常那样）我确信真理在我这边，却无法反驳他们。

我计划9月23日星期二飞回美国。星期一晚上，我正在大学招待所的小屋里收拾行李，电话铃响了，是泽尔多维奇的：“到我这儿来，基普！我想和你谈谈黑洞的蒸发！”时间很紧，我想在门口找辆的士，但一个也没有。于是我像地道的莫斯科人那样，拦了辆摩托，给他5卢布，让他送我到绍瑟街2B号。他答应了。摩托开上一条我从没走过的偏僻街道，我真怕走丢了。等转弯上了绍瑟街，我才放下心

来。说声“谢谢”，我在2B门口下了车，轻轻穿过铁门和树荫，走进一座楼，登上楼梯，走向二楼西南角。

泽尔多维奇和斯塔罗宾斯基在门口等我，满脸带笑，双手举过头。“我们投降了。霍金是对的，我们错了！”在接下来的几个小时里，他们对我说，他们的黑洞弯曲时空的量子场，虽然在形式上看起来与霍金的不同，但实际上是完全等价的。他们原来讲黑洞不能蒸发是因为计算出了错，不是定律的问题。现在错误纠正了，他们也同意，定律要求黑洞蒸发，没有什么例外。

可以用几种不同的方式来描绘黑洞的蒸发，它们相应于以不同的方法建立黑洞弯曲时空的量子场定律。然而，所有方法都认为真空涨落是向外辐射的最终源泉。最简单的描绘也许是建立在粒子（而不是波动）基础上的：[19]

真空涨落跟“真”的正能量波一样服从波粒二象性定律（卡片4.1），也就是，它们有波的一面，也有粒子的一面。波的那一面我们已经见过了（卡片12.4）：波随机而不可预测地涨落，一会儿在这里是正能量，一会儿在那里是负能量，平均起来却没有能量。粒子的一面体现在虚粒子的概念上，就是说粒子凭着从邻近空间区域借来的涨落能量成对地闪现（同时出现两个粒子），然后湮灭而消失，把能量还给邻近的区域。对电磁真空涨落来说，虚粒子是虚光子；对引力真空涨落来说，是虚引力子。[2]

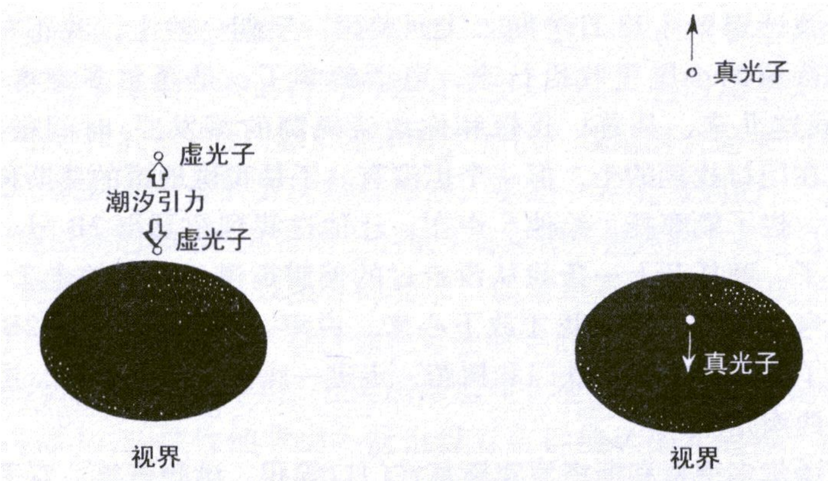



图12.2　落向黑洞的观察者所看到的黑洞蒸发机制。左：黑洞潮汐引力将一对虚光子分开，从而向它们提供能量。右：虚光子从潮汐引力获得足够能量而暂时物质化为真实光子，一个离开黑洞，而另一个落进黑洞中心。

图12.2描绘了真空涨落是如何导致黑洞蒸发的。左边画的是某个落向黑洞的参照系所看到的黑洞视界附近的一对虚光子。虚光子对很容易分开，只要它们所在区域的电磁场瞬间获得正能量。那个区域可以很小，也可以很大，因为真空涨落在一切波长尺度上都可能发生。但区域大小总是大致和它涨落的电磁波的波长相同，所以虚光子只能分开约一个波长。如果波长正巧和黑洞的周长一样，那么虚光子很容易像图上那样分开四分之一周长。

视界附近的潮汐引力很强；在光子之间下落的观察者看来，强大的力量将虚光子分开，从而也向它注入巨大能量。虚光子因能量的增加，到它们分离四分之一视界周长时，就足以转化为实在的长寿命光子（图12.2右），也有足够能量留下来还给相邻的负能的空间区域。现在的实光子相互解脱了，一个在视界内部，从外面的宇宙中永远消失；另一个脱离黑洞，带走了潮汐引力给它的能量（也就是物质[3]）。黑洞因失去质量而有一点收缩。

这种粒子发射机制并不依赖于粒子是否是光子，相关的波是否是

电磁波。它对所有其他形式的粒子和波（也就是对所有其他类型的辐射——引力的、中微子的等等）也同样适用，因此黑洞会产生所有类型的辐射。

虚粒子在物质化为实粒子前必然靠得较近，距离大概小于它们的波长。然而，为了从黑洞潮汐引力得到足够物质化的能量，它们必须分离约黑洞周长的四分之一。这意味着，黑洞发射的波或粒子的波长约为黑洞周长的四分之一，或者更大。

两个太阳质量的黑洞周长约为35千米，所以它发射的粒子或波的波长约为9公里或更大。同光或普通无线电波相比，这是巨大的波长，但与两个黑洞碰撞时可能发射的引力波的波长相比，它并没有多大差别。

霍金在刚从事研究的那些年，总是想做到非常仔细、非常严格。在事情没有得到几乎无懈可击的证明前，他从来不说它是对的。然而，到1974年，他的态度变了："我更愿意正确，而不是严格。"他曾这么坚决地告诉我。达到高度的严密需要花费很多时间。这一年，霍金为自己定下了两个目标：认识广义相对论与量子力学的完全结合，认识宇宙的起源——实现这些目标，需要大量的时间和精力。也许因为自己那要命的病，霍金比别人更能体会生命的有限。他觉得他不可能为了达到高度严密而长时间地停在他的发现上，也没有精力去探索那些发现的所有重要特征。他必须尽快向前赶。

于是，我们看到，霍金在1974年严格证明了黑洞像一个具有正比于其表面引力的温度的热物体那样辐射后，在没有真正证明的情况下又接着宣称，黑洞力学定律与热力学定律的所有其他相似也都不是简单的巧合：黑洞定律与热力学定律是同样的东西，不过外表不同罢了。根据这一论断和他严格证明的温度和表面引力之间的关系，霍金猜测了黑洞的熵和它的表面积之间的精确关系：熵是0.10857...乘以表面积，除以普朗克-惠勒面积。[4]换句话说，10个太阳质量的非旋转

黑洞具有的熵是$4.6 \times 10^{78}$，近似于贝肯斯坦的猜想。

贝肯斯坦当然相信霍金是正确的，他感觉很满意。1975年底，泽尔多维奇、斯塔罗宾斯基、我和霍金的其他同事也非常愿意同意他的观点。然而我们并不完全感到满意，因为还没认识到黑洞的巨大随机性的本质。黑洞内部的某些东西一定有$10^{4.6 \times 10^{78}}$种分布方式，而这些分布却不会改变它的外在表现（质量、角动量和电荷），它们是些什么东西呢？另外，我们如何能够通过简单的物理学关系来认识黑洞的热行为——也就是黑洞行为像一个具有一定温度的普通物体这一事实呢？霍金继续向前去研究量子引力和宇宙起源，戴维斯（Paul Davies）、昂鲁什、瓦尔德、约克、我和他的许多同事则瞄准了这些问题。在未来的10年里，我们逐步获得了一些新认识，表现在图12.3中。[20]

图12.3（a）表现的是下落经过视界的观察者所看到的黑洞的真空涨落。真空涨落由对对虚粒子构成。潮汐引力偶尔给这么多粒子对中的某一对以充足的能量，使它的两个虚粒子成为实在的，然后其中一个脱离黑洞。这是图12.2讨论过的真空涨落和黑洞蒸发的观点。

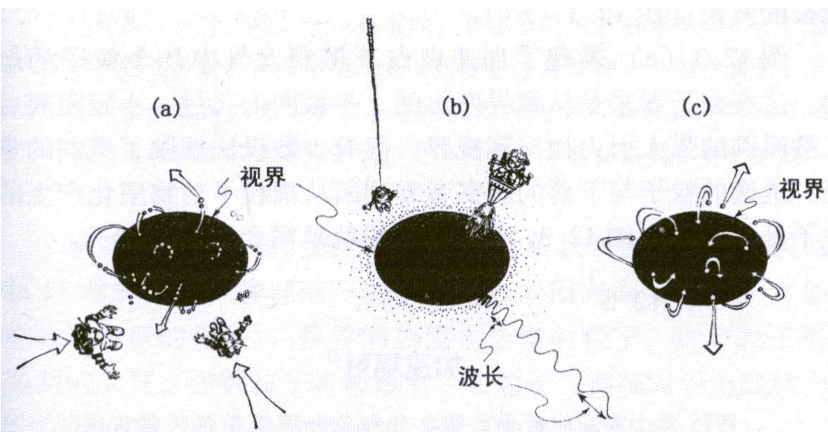



图12.3　（a）落进黑洞的观察者（穿太空服的两个小人）看到黑洞视界附近的真空涨落由虚粒子对构成。（b）在视界上方相对静止的观察者（绳子吊着的小人和点燃火箭的小人）看

来，真空涨落由真实粒子的热大气组成，这是“加速的观点”。（c）以加速的观点看，大气粒子似乎是从热的膜状视界发出来的。它们向上飞过一小段距离，然后多数被拉回视界，然而还有少数粒子设法逃脱了黑洞的掌握，蒸发进入外面的空间。

图12.3（b）描述了一个不同的黑洞真空涨落的观点，停在视界上方并永远相对于它静止的观察者的观点。这样的观察者相对于下落的观察者一定要艰难地加速向上，才不致被黑洞吞没——他靠火箭的反冲或者用绳子吊起来。因为这个理由，这些观察者的观点叫做“加速的观点”，它也是“膜规范”的观点（第11章）。

奇怪的是，从加速的观点看，真空涨落不是飘忽出没的虚粒子，而是具有正能量和长寿命的真粒子，见卡片12.5。真粒子在黑洞周围形成像太阳大气那样的热气。与这些真粒子相联系的是真实的波。当粒子向上运动穿过大气时，引力的作用将减小它的动能；相应地，当波向外传播时，会因引力作用而红移到越来越长的波长（图12.3（b））。

图12.3（c）表现了加速观点下黑洞大气中几个粒子的运动。这些粒子看来是从视界发出的，多数向上飞过一小段距离后又被黑洞的强大引力拉回到视界，但有少数设法摆脱了黑洞的掌握。逃逸的粒子与下落的观察者看到的从虚粒子对物质化产生的粒子是一样的（图12.3（a）），它们就是霍金的蒸发粒子。

卡片12.5　加速辐射[21]

1975年，惠勒的新学生昂鲁什和伦敦国王学院的戴维斯（用弯曲时空的量子场定律）独立发现，黑洞视界上方的加速观察者一定会看到那里的真空涨落不是虚粒子对，而是真实粒子的大气。昂鲁什把这种大气叫“加速辐射”。

这个惊人的发现揭示了真粒子的概念是相对的，而不是绝对的；就是说，它依赖于观察者的参照系。在自由下落的参照系中进入黑洞视界的观察者看不到视界外的真粒子，只能看到虚粒子。加速参照系

中的观察者靠自己的加速度而总留在视界上方，能看到许许多多真实的粒子。

这怎么可能呢？一个观察者称视界被真实粒子的大气包围着，而另一个观察者却说不是那样的，能有这样的事吗？答案在于这样一个事实：虚粒子的真空涨落波并不严格限于视界外的某个区域，每一涨落波都是部分在视界内，部分在视界外。

·自由下落穿过视界的观察者能看到真空涨落波的两个部分，即在视界外的和在视界内的；所以，这样的观察者很清楚（凭他们的测量），波只是真空涨落，它相应的粒子是虚的，而不是实的。

·留在视界外面的加速观察者只能看到真空涨落波的外面部分，看不到它在视界内的部分；这样，根据他们的测量，他们不能判别波只是伴随虚粒子的真空涨落。因为只看到了涨落波的一部分，他们就误认为它是“真实的东西”——伴随真实粒子的真实波动，结果，他们的测量表明视界周围是一片真实粒子的大气。这种真实粒子的大气会逐渐蒸发，飞向外面的宇宙（图12.3（c））。

这个事实表明，加速观察者的观点与自由下落观察者的观点实际上是一样的，一样正确，也一样有效：自由下落的观察者看到的是，虚粒子对在潮汐引力作用下转化为真粒子，然后其中一个粒子蒸发；而加速观察者看到的更简单，总是在黑洞大气中的永远真实的粒子蒸发了。两种观点都是对的；它们是不同参照系看到的同一物理景象。

从加速的观点看，视界就像高温的膜状表面，这里的膜就是第11章所说的“膜规范”的膜。正如太阳表面发出粒子（如照亮地球白昼的光子），视界的热膜也会发射粒子，这些粒子形成黑洞的大气，少数粒子将被蒸发。在粒子飞离膜时引力红移会减少它的能量，所以虽然膜很热，蒸发的辐射却很冷。

加速观点不仅解释了黑洞在什么意义上是热的，而且还说明了黑

洞巨大的随机性。下面（由我和我的博士后朱里克（Wo-jciech Zurek））设计的思想实验解释了那是怎么回事。

向黑洞大气投入少量能量（或质量）、角动量（旋转）和电荷。这些东西将从大气向下穿过视界进入黑洞。一旦注入物质进去了，就不可能从黑洞外面了解它们的性质（是物质的，还是反物质的；是没有质量的光子还是有质量的原子；是电子还是正电子），也不可能知道它们是从哪儿来的。因为黑洞无“毛”，我们通过黑洞外的考察所能了解的，只是进入大气的总的质量、角动量和电荷。

那么，这些质量、角动量和电荷能以多少种方式注入黑洞的热大气呢？这个问题类似于问在卡片12.3中的玩具屋里，孩子的玩具能有多少种方法堆放在地砖上。相应地，注入方式的总数的对数，如标准热力学定律所说的，必然就是黑洞大气熵的增量。朱里克和我通过很简单的计算就证明了，这个增加的热力学熵正好等于增加的视界面积的1/4，除以普朗克-惠勒面积；也就是说，它实际上是另一种形式的面积增加，与霍金在1974年根据黑洞力学定律与热力学定律的数学相似性所猜测的一样。这个思想实验的结果还可以像下面这样说得更简洁些：黑洞的熵是能形成它的方式的数目的对数。这意味着，有$10^{4.6\times10^{78}}$种不同方法可以形成一个熵为$4.6\times10^{78}$的10个太阳质量的黑洞。这个熵的解释原来是贝肯斯坦在1972年猜想的，霍金和他以前的学生吉本斯（Gary Gibbons）1977年给出了一种高度抽象的证明。[22]

这个思想实验也证明热力学第二定律仍然发挥着作用。投进黑洞大气的能量、角动量和电荷可以是任意形式的，例如，可以把满屋的空气装进一个袋子，前面我们考虑第二定律时已经见过了。当袋子投入黑洞大气时，外面宇宙的熵将减少袋子所具有的熵（随机性）。然而，黑洞大气的熵，从而黑洞的熵，却增加得更多，所以黑洞的熵加上外面宇宙的熵的总和还是增加了，服从热力学第二定律。

同样，我们会看到，黑洞在蒸发了一些粒子后，自己的表面积和熵通常会下降，但粒子在外面宇宙的随机分布增加了宇宙的熵，大大超过了黑洞失去的熵。这样，第二定律仍然是满足的。

黑洞蒸发和消失需经历多长时间呢？答案依赖于黑洞的质量。黑洞越大，温度越低，于是发射粒子越弱，蒸发也就越慢。1975年，当帕奇还是我和霍金的学生时，曾做过计算，[23]假如黑洞质量是太阳的两倍，那么它的寿命是$1.2\times10^{67}$年。黑洞寿命正比于质量的立方，所以，20个太阳质量的黑洞的寿命为$1.2\times10^{70}$年。这些年龄同宇宙目前$1\times10^{10}$年的年龄相比，真是太大了，所以蒸发不会影响天体物理学。不过，对我们认识广义相对论与量子力学的结合来说，蒸发还是很重要的。我们从认识蒸发的努力中学会了弯曲时空的量子场。

质量远小于2个太阳的黑洞如果存在的话，蒸发起来远远不会像$10^{67}$年那么漫长。这样的小黑洞在今天的宇宙中是不会形成的，因为物质的简并压力和核压力很强，即使当今宇宙的一切力量都来挤压它们，这些物质也不会坍缩（第4，5章）。然而，在宇宙大爆炸时可能会产生这样的黑洞，[5]那时物质所经历的密度、压力和引力挤压都远远高过现在的恒星。

霍金、泽尔多维奇、诺维科夫和其他一些人的详细计算表明，从大爆炸出来的物质小集团可以产生小黑洞，[24]只要这些成团物质的状态方程是“软”的（也就是在挤压时只增加很小的压力）。在极早期的宇宙中，相邻物质像把强力砧板上的碳挤压成金刚石那样，也把那些小集团挤压成小黑洞。

寻找那些原生小黑洞的一个有希望的办法，是寻找它们蒸发产生的粒子。质量小于5 000亿千克（$5\times10^{14}$克，一座不太大的山的重量）的黑洞到现在可能刚蒸发完，比它重几倍的黑洞现在应该在剧烈蒸发中。这些黑洞的视界大约是一个原子核的大小。

从这些黑洞的蒸发中发出的能量现在大部分应该表现为在宇宙中随机穿行的γ射线（高能光子）。这样的γ射线确实存在，但它的数量和性质很容易用别的方式来解释。（根据霍金和帕奇的计算，）没有多余的γ射线的事实告诉我们，现在在每立方光年的空间里，强烈蒸发的小黑洞不会多于300个，这也就告诉我们，大爆炸时的物质不可能有特别软的状态方程。[25]

怀疑者会说，为什么没有多余的γ射线，可能有另一种解释：也许大爆炸中形成过许多小黑洞，但我们物理学家对弯曲时空量子场的认识远不像我们想象的那么好，所以，当我们相信黑洞蒸发时，正在走向错误的方向。我和我的同事不同意这种怀疑，因为我们看到，标准的弯曲时空定律和标准的量子场定律完美地融合在一起了，为我们带来了几乎惟一的一组弯曲时空的量子场论定律。不过，如果天文学家能找到黑洞蒸发的观测证据，我们会更满意的。

[1] 我们说“几乎”，是考虑了用来计算真空涨落能量的所谓“重整化”过程有一定的模糊性，这些由瓦尔德（惠勒以前的学生）发现和分析过的不确定性不影响黑洞的蒸发，不过也可能只有在掌握了引力的完全的量子理论后才会得到解决。[18]

[2] 可能有些读者已经在物质和反物质背景下熟悉这些概念了，例如，一个电子（物质粒子）和一个正电子（反粒子）。像电磁场是光子的场一样，也存在电子和正电子的场，即电子场。在电子场真空涨落瞬时巨大的地方，虚电子和虚正电子很可能成对出现，当场衰落时，电子和正电子也很容易湮灭而消失。光子是自身的反粒子，所以虚光子也成对地出现和消失，引力子也一样。

[3] 回想一下，由于质量和能量完全可以相互转化，所以它们实际上只是同一概念的不同名称。

[4] 这个特别因子0.10857…实际上是1/（4$\log_e$10），这里$\log_e$10、＝2.30258…来自我选择的熵的“正规化”。

[5] 粗略地说，黑洞的质量是宇宙形成的时间（秒）乘以$10^{38}$（克），例如，在普朗克-惠勒时间，原生小黑洞质量约为10毫克，而在万分之一（$10^{-4}$）秒时，黑洞就有太阳那么重了。——译者

# 第13章　洞里

同爱因斯坦场方程搏斗的物理学家

寻找黑洞里的秘密：

一条通向其他宇宙的道路？

一个无限潮汐引力的奇点？

是空间和时间的终结，还是量子泡沫的源泉？

## 奇点和其他宇宙

黑洞里面是什么？

我们如何去认识，我们为什么关心？没有信号能从洞里出来告诉我们答案。没有哪个勇敢的探险家在走进黑洞弄清楚后还能走出来告诉我们，他甚至向我们发回答案也做不到。不论黑洞中心是些什么，它们都不可能出来以任何方式影响我们的宇宙。

人的好奇心是不会满意这种回答的。何况，我们还有可以告诉我们答案的工具：物理学定律。

惠勒曾向我们讲过认识黑洞中心的重要性。他在50年代提出引力坍缩的“最终状态的问题”，说这是理论物理学家们的圣杯，[1]它可能为我们彻底揭示广义相对论与量子力学的“火热结合”。在奥本海默认为最终状态隐藏在视界的背后时，惠勒就反对（第6章）——我想，主

要原因还是他为不能从视界外面看到那火热结合的场面而感到痛苦。[1]

惠勒承认有那么一个视界，不过他还是坚持认为黑洞的中心是值得追求的圣杯。[2]对黑洞蒸发的理解曾帮助我们实现了量子力学与广义相对论的部分结合（第12章），那么对黑洞中心的认识也许会让我们发现它们完全的结合，为我们带来一套完全的量子引力定律。中心也许还藏着打开宇宙其他奥秘的钥匙：在亿万年后我们宇宙死亡时可能的"大挤压"坍缩与恒星生成黑洞时的坍缩之间存在着一种相似，我们把握了一个，就能认识另一个。

惠勒的圣杯，物理学家已追寻了35年，成绩却很小。我们还不知道黑洞中心有什么，认识的努力还没能带来清晰的量子引力。不过，我们也学会了不少——特别是，不论黑洞里的东西是什么，都与量子引力定律紧密联系着。

这一章要讲的是我们在追寻惠勒圣杯的路上的一些更有趣的崎岖经历和转折以及我们目前所到达的地方。

"黑洞里有什么？"第一个答案来自奥本海默和斯尼德1939年关于球状恒星坍缩的经典计算（第6章）。[3]虽然答案包含在他们发表的方程里，但他们并没有讨论。也许，他们害怕火上浇油，因为人们正在争论他们关于坍缩恒星"把自己同宇宙隔绝开来"（也就是形成一个黑洞）的预言；也许因为奥本海默天生的科学保守态度，他不愿猜想，所以保持沉默。[4]不管怎样，他们什么也没说，但那方程把什么都说了。

他们的方程说，球状恒星生成包围自己的黑洞视界后会继续无情地坍缩下去，直到没有体积却有无限大的密度，它在这里产生时空奇点，也在奇点处消失。

依照广义相对论，奇点是时空曲率无限大因而时空不再存在的区

域。由于潮汐引力是时空曲率的表现（第2章），所以，奇点也是潮汐引力无限大的区域。就是说，那里的引力将在某些方向无限拉伸，而在其他方向无限挤压所有的物体。

我们可以想象不同种类的时空奇点，各自具有特别的潮汐涨落特征，在这一章里我们会遇到几种。

奥本海默-斯尼德计算所预言的奇点是最简单的一类。[5]它的潮汐引力基本上与地球、月亮和太阳的相似，也就是与在地球上产生海洋潮汐的力一样（卡片2.5）：奇点在径向上（朝向或背离它的方向）拉伸所有物体，在横向上挤压所有物体。

假设有个落向奥本海默和斯尼德的方程所描绘的那类黑洞的宇航员。黑洞越大，他活得越久，为让他多活些时候，假定黑洞是类星体中最大的那种（第9章）：100亿个太阳那么重。下落的宇航员通过视界，在临死前20小时进入黑洞。不过他刚进来时还离奇点太远，感觉不到它的潮汐引力。随着越落越快，离奇点越来越近，潮汐引力变得越来越强，在离奇点1秒前，他开始觉得它在将他的头和脚分开，而且从两肋将他挤压（图13.1底）。起初，他对这样的拉挤还不感到很痛，但力量继续增大，在他到奇点的一百分之几秒前（中图），他的骨肉就抵抗不住了，身体分离，死了。在最后百分之一秒内，拉伸和挤压还在增强，当他到达奇点时，那些力量已经无限大了。先作用在他脚上，然后作用到躯干和头颅，身体被无限拉长，最后，根据广义相对论，他成为奇点的一部分，消失了。

宇航员绝对不可能继续穿过奇点而从另一边出来，因为广义相对论认为没有“另一边”。在奇点处，空间、时间和时空都不存在。奇点是一个分明的边缘，像一张纸的边缘一样。纸的边缘外没有纸，奇点外也没有时空。不同的是，纸上的妈蚁可以爬到边缘，然后退回来，但任何东西都不可能离开奇点；根据爱因斯坦的广义相对论定律，不管是宇航员、粒子、波还是别的什么，只要碰到奇点，都会在瞬间毁

灭。

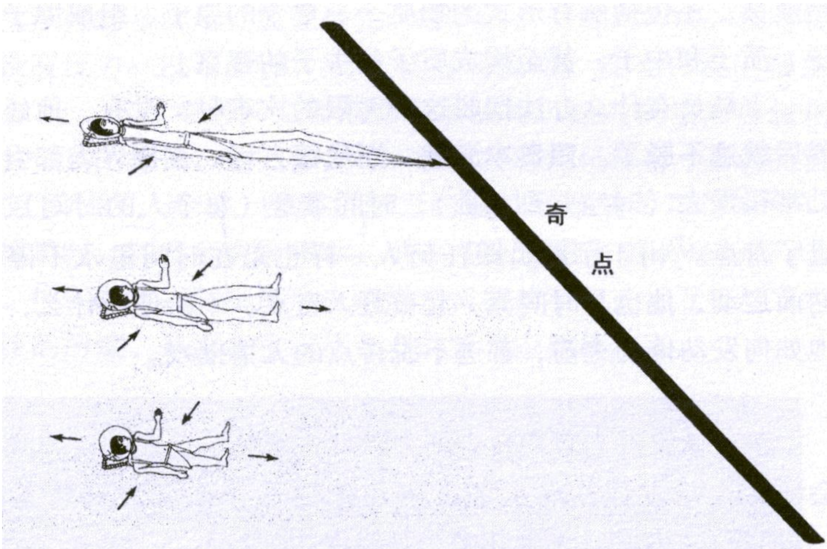



图13.1　根据奥本海默-斯尼德计算，一个宇航员（脚先着地）落进黑洞中心奇点的时空图。像以前所有的时空图一样（例如，图6.7），少画了一个空间维，所以宇航员看起来是2维的而不是3维的。奇点在本图中是斜的，而不像在图6.7和卡片12.1中那样是垂直的，这是因为在这里向上的时间和水平的空间与别处不同，它们是宇航员自己的时间和空间，而在其他地方则是芬克尔斯坦的。[6]

图13-1并没有将破坏机制完全表达清楚，因为它忽略了空间的曲率。事实上，宇航员在到达奇点时真被拉得无限长，而在横向上被压得没影儿了。奇点附近的这种极端空间曲率能令他无限伸长却不能将头钻出黑洞视界。他的头和脚都被拉进了奇点，却分离无限远。

照奥本海默和斯尼德的方程，受无限拉伸和挤压的不仅只有宇航员，还包括所有形式的物质——单个的原子、组成原子的电子、质子和中子，甚至构成质子和中子的夸克。

宇航员有什么办法摆脱这种无限的灾难吗？没有。他经过视界后就逃不脱了。照奥本海默-斯尼德方程，在视界内部引力到处都很强大（时空强烈卷曲），时间本身（每个人的时间）也流进了奇点。[2]由

于宇航员跟任何人一样也是在时间里永不停歇地向前运动，他也与时间流一起被赶入奇点。不论他做什么，不论他如何发动他的火箭，都逃不脱奇点的无情摧残。

物理学家每当看到我们的方程预言了某些无限的东西时，总会怀疑这些方程。现实宇宙中几乎没有什么东西真是无限的（我们想），因此，无限几乎总是错误的信号。

奇点的无限拉伸和挤压作用也不会例外。50年代和60年代初研究过奥本海默和斯尼德论文的那些物理学家都一致认为，一定在哪儿出了错，但分歧也跟着来了。

在惠勒强有力的领导下的一个研究小组，认定这种无限作用确凿说明广义相对论在黑洞内部星体坍缩的终点失败了。[7]惠勒断言量子力学能阻止那里的潮汐引力变得真正无限大；但是它怎么做呢？惠勒说，为得到答案，我们需要把量子力学的定律与潮汐引力的定律，也就是爱因斯坦广义相对论的弯曲时空定律结合起来。惠勒宣称，结合的产儿量子引力定律，一定会征服奇点；而且新定律还可能产生黑洞内部的一些新物理现象，与我们以前遇见的都不一样。

卡拉特尼科夫（Isaac Markovich Khalatnikov）和栗弗席兹（Ergeny Michailovich Lifshitz，朗道的莫斯科小组成员）领导的另一组则认为，无限的作用警告我们，奥本海默和斯尼德的理想化坍缩恒星模型是不可信的。[8]回想一下，奥本海默和斯尼德的计算所依赖的基础是：恒星是完全球状的，密度均匀的，没有旋转，没有压力，没有激波，没有喷射的物质，没有外流的辐射（图13.2）。卡拉特尼科夫和栗弗席兹认为，这些极端的理想化是产生奇点的根源。他们称，每一颗真实的恒星都有小小的随机的形变（形状、速度、密度和压力的随机的微弱不均匀变化），恒星坍缩时，这些形变会增长，并在奇点形成之前使坍缩停止。同样，他们断言，随机形变也将阻止我们的宇宙在亿万年后发生大挤压的坍缩，从而将宇宙从奇点的毁灭中

拯救出来。

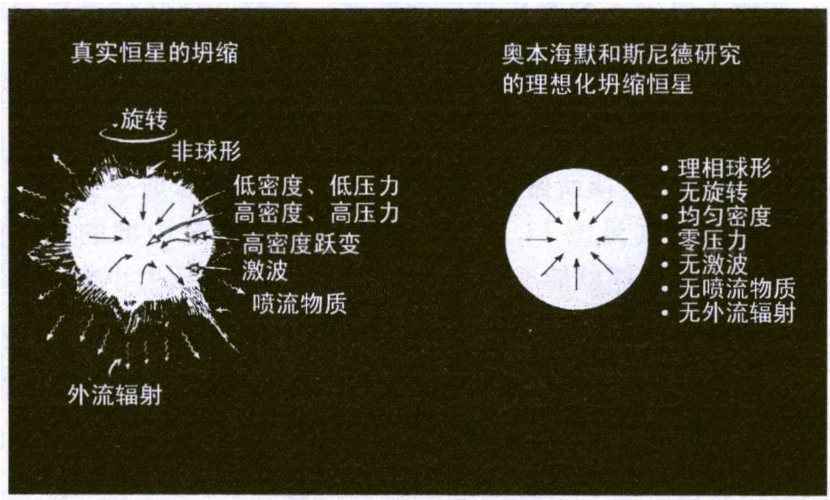



图13.2　（同图6.3）左：现实的坍缩恒星的物理现象。右：奥本海默和斯尼德为计算星体坍缩而做的理想化。具体讨论见第6章。

卡拉特尼科夫和栗弗席兹的这些观点来自他们1961年向自己提出的一个问题：照爱因斯坦的广义相对论定律，奇点对小的扰动是否是稳定的？[9]换句话讲，他们针对奇点提的问题，与我们在第7章里遇到的一个关于黑洞的问题是一样的：假如在解爱因斯坦场方程时我们以微小但随机的方式改变坍缩恒星或宇宙的形状，改变物质的速度、密度和压力，并向这些物质注入少量随机的引力辐射，那么，这些改变（微扰）会给预言的坍缩结果带来什么影响呢？

我们在第7章已经看到，对黑洞视界而言，这些扰动不会带来什么影响。被扰动的坍缩的恒星仍然形成视界，尽管视界开始有些变形，但所有变形很快会辐射开去，留下一个完全“无毛”的黑洞。也就是说，视界对小扰动是稳定的。

但是，对黑洞中心或宇宙最后挤压的奇点来说，卡拉特尼科夫和栗弗席兹得到了与别人不同的结论。他们的计算似乎说明，在坍缩物

质形成奇点的过程中，小小的随机扰动会长大，多大呢？实际上可以大到阻止奇点的形成。大概（尽管计算还不能说肯定）这些扰动将阻止坍缩而使它爆炸。

微扰怎么可能扭转坍缩呢？在卡拉特尼科夫-栗弗席兹计算中，物理机制还一点儿没弄清楚。不过，用牛顿引力定律所做的（比用爱因斯坦定律所做的）简单得多的其他计算似乎提供了一些线索。例如（见图13.3），假如一颗坍缩恒星内部的引力很弱，这样牛顿定律就能准确应用；又假如恒星的压力小得无足轻重，那么小扰动将使不同原子落向恒星中心附近不同的地方。大多数坍缩的原子将偏离中心一定的距离，并且绕着中心摆动，然后飞出去，这样坍缩就转变成了爆炸。可以想象，即使牛顿引力定律不适用于黑洞内部，某个类似的机制也可能将坍缩逆转为爆炸。

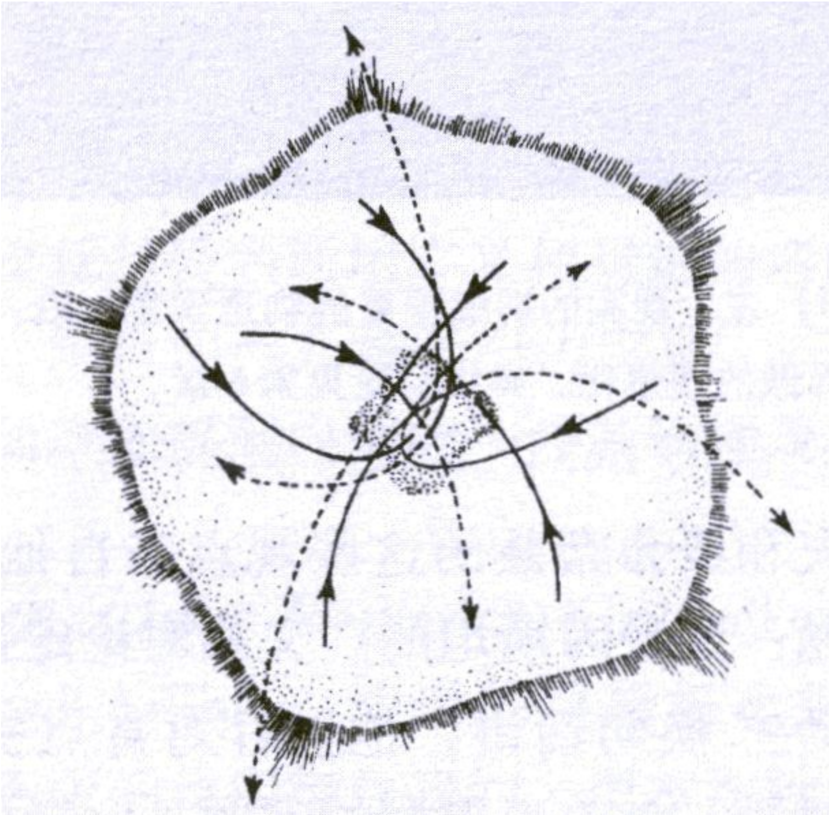


图13.3　在引力微弱、牛顿定律精确适用和内部压力无关紧要的条件下，恒星坍缩转变为爆炸的一种机制。假如坍缩的恒星有一点变形（“扰动”），它的原子会落到不同的位置，绕着中心摆动，然后飞出去。

我是1962年作为研究生加入惠勒的研究小组的，那时卡拉特尼科夫和栗弗席兹刚发表他们的计算，栗弗席兹和朗道也刚把计算和“无奇点”的结论写进一本有名的教科书[10]《经典场论》。[3]我清楚记得，惠勒鼓励大家去研究这些计算，他告诉我们，如果它们对了，结果是深刻的。不幸的是，计算又长又复杂，而发表的细节又太简略，我们很难检验——而卡拉特尼科夫和栗弗席兹正被罩在苏联的铁幕下，我们不能坐到一起来详细讨论。

不过，我们还是开始考虑这样一种可能：坍缩的宇宙在到达某个

很小的尺度时，也许会“反弹”，再发生新一次“大爆炸”；同样，坍缩的恒星收缩到视界内部后，也可能反弹并再次爆炸。

但是，假如恒星再爆炸了，它会成什么呢？它当然不可能爆出黑洞的视界。爱因斯坦的引力定律严禁任何事物（虚粒子例外）飞出视界。不过还有别的可能，恒星爆炸后，可能进入我们宇宙的某些其他区域，甚至进入别的宇宙。

图13.4用一个嵌入图序列描绘了这样一个从坍缩到再爆炸的过程。（嵌入图与时空图大不一样，是我们在图3.2和3.3引入的。）

每一幅图都将我们宇宙的弯曲空间和别的宇宙的弯曲空间画成嵌在高维超空间里的二维曲面。（回想一下，超空间是物理学家想象的东西：人类生活总是被限制在我们的宇宙空间（或别的宇宙空间，如果我们能去的话）；我们永远不可能走出这个空间而进入周围更高维的超空间，也不可能收到来自超空间的信号或信息。超空间不过是形象化的工具，例如，它帮助我们形象地表现枬缩恒星和它的黑洞周围的空间如何弯曲，恒星如何在我们的宇宙中坍缩然后又爆炸而进入别的宇宙。）

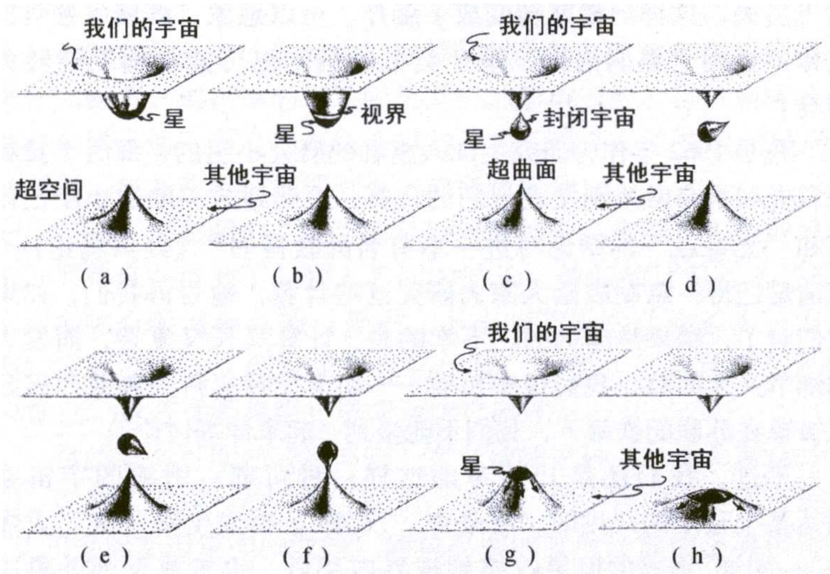



图13.4 坍缩成黑洞的恒星的一种可能的命运（不过，在本章后面会看到，那是很不可能的）的嵌入示意图。这8幅图（a）到（h）是表现恒星和空间几何演化的一个快照序列。恒星在我们宇宙中坍缩（a），形成包围它的黑洞视界（b）。然后，黑洞深处的空间区域内包含着一个脱离我们宇宙的一小团星体，形成一个与任何事物都没有联系的封闭小宇宙（c）。接着，封闭小宇宙在超空间中运动（d，e），将自己同另一个宇宙相联（f）；然后星体爆炸，进入那个宇宙（g，h）。[11]

在图13.4中，两个宇宙像海洋里两个分开的岛，超空间就像海洋的水。岛是没有大陆联结的，两个宇宙也不存在空间联系。

图13.4的序列描绘了恒星的演化。在我们的宇宙中，恒星开始坍缩（图a）），在图（b）它已经形成了包围自己的视界，而且继续坍缩。在图（C）和（d），恒星被高度压缩的物质使周围空间强烈弯曲，形成一个封闭的像气球表面的小宇宙；这个新的小宇宙从我们的宇宙中挤落下来，孤单地进入超空间。（这有点儿像一个岛上的居民造一只小船，划着它过海。）在图（d）和图（e），小宇宙带着恒星内部从我们的大宇宙穿越超空间到达另一个大宇宙（小船从一个岛划到另一个岛）。在图（f），小宇宙同那个大宇宙连起来（小船登岸），继续膨胀，吐出恒星。在图（g）和图（h），恒星爆炸，进入

那个宇宙。

这样听来纯乎是科幻小说的画外音，是不能令我满意的。然而，正如黑洞是爱因斯坦场方程的史瓦西解的自然结果（第3章），这幅图景也是爱因斯坦方程另一个解的自然产物，那是雷斯纳和诺德斯特勒姆在1916～1918年间发现的，但他们并没完全理解。1960年，惠勒的两个学生布雷尔和格雷弗斯（John Graves）找到了雷斯纳-诺德斯特勒姆解的物理意义[12]，很快大家就明白，不必太多的改变，这个解就能描绘图13.4的恒星坍缩和爆炸。这颗恒星与奥本海默和斯尼德的恒星正好存在一个根本性的差别：它自身携带的电荷足以在高度致密时产生很大电场，这个电场似乎以某种方式影响着恒星再次爆炸进入另一个宇宙。

让我们来看看，1964年在追寻惠勒的圣杯——也就是认识坍缩成黑洞的恒星的最终命运——的过程中，我们发现了些什么。

1.我们知道，爱因斯坦方程有一个解（奥本海默-斯尼德解）预言，假如恒星具有高度理想的形态（包括理想的球形），则它会在黑洞中心产生一个有无限潮汐引力的奇点——一个捕获、破坏并吞噬一切进入黑洞的事物的奇点。

2.我们知道，爱因斯坦方程还有一个解（雷斯纳-诺德斯特勒姆解的推广）预言，假如恒星有某些不同的高度理想化形态（包括球形和电荷），则在黑洞内部深处，恒星将从我们的宇宙脱落，与别的宇宙相连（或者到一个远离我们宇宙的区域），并在那儿再发生爆炸。

3.我们还远没弄清楚，这些解中哪个“对小的随机扰动是稳定的”，从而可能在真实宇宙中发生。

4.然而，卡拉特尼科夫和栗弗席兹声称他们证明了奇点对小的随机扰动总是不稳定的，所以永远不可能出现，从而奥本海默-斯尼德奇

点也永不可能出现在我们真实的宇宙中。

5.至少，普林斯顿有人怀疑卡拉特尼科夫和栗弗席兹的证明。这些怀疑可能部分是由于惠勒对奇点的需要引起的，因为奇点可能是广义相对论和量子力学“结合的殿堂”。

1964年是转折的一年。在这一年，彭罗斯为我们带来了革命性的用以分析时空性质的数学工具。他的革命太重要了，极大地影响着我们对惠勒圣杯的追求。所以，我要拿出几页来讲他的革命和他这个人。

[1] 圣杯（Holy Grail）是耶稣基督在最后的晚餐上用的酒杯，“这杯是用我血所立的新约，那血是为你们流的”（《新约·路加福音》）。现在多用来比喻一个人终生追求的目标。——译者

[2] 用专业术语，我们说奇点是“类空的”。

[3] 这本书的汉译《场论》是根据1948年版俄文本翻译的，没有这里和后面谈到的内容。——译者

# 彭罗斯的革命

罗杰·彭罗斯生在一个医学家庭，[13]母亲是医生，父亲是伦敦大学学院[1]的知名人类遗传学教授。父母希望四个孩子中至少有一个跟他们一样从医。罗杰的哥哥奥利弗（Oliver）是不会学医的，他很小就想学物理（后来真成了世界有名的统计物理学家——研究大量相互作用的原子的行为）。弟弟约拿坦（Jonathan）也不会，他只想下棋（后来连续6年成为英国的象棋冠军）。妹妹雪莱（Shirley）还小，在罗杰选择职业的时候，她还没有什么倾向的表现（她长大后终于成了医生，满足了父母的心愿）。剩下罗杰是父母最大的希望了。

16岁时，罗杰和班上别的同学一样，被校长找去谈话。他现在该决定上大学前下两年的学习科目。他告诉校长，“我想学数学、化学和生物学。”校长声明，“不行，那不可能。生物和数学不能联在一起，你只能选一门。”罗杰更喜欢数学，不太喜欢生物学。“那好，我学数学、化学和物理。”他说。那天晚上罗杰回家后，父母很生气。他们训斥罗杰跟伙伴学坏了。要做医生，生物学是最基本的，他怎么能放弃呢？

两年后，该决定上大学学什么了。罗杰回忆说，“我提出到伦敦大学学院学数学，父亲完全不同意。他说，对做不了其他事情的人来说，数学也许是恰当的，但把它作为实际的职业，就不对了。”罗杰一定要学，父亲于是请学院的一个数学家来特别考考他。数学家要罗杰用一整天来考试，警告他说，可能只解得了一两个问题。罗杰只用了几个小时就答对了全部12道题，父亲同意了，罗杰可以学数学。

罗杰原来没想过要把数学用到物理上来，他只对纯数学有兴趣，但他经不住诱惑。

诱惑是从1952年开始来的。[14]那时罗杰是伦敦的大学四年级学

生，在广播里听了霍伊尔（Fred　Hoyle）的一系列宇宙学讲座。讲座有趣、动人——也有一点儿难以理解。霍伊尔说的某些东西简直没有意义。一天，罗杰坐火车到剑桥去看在那儿学物理的哥哥奥利弗。晚餐时，在金斯伍德餐厅，罗杰发现跟奥利弗一个办公室的席艾玛正在研究邦迪-戈尔德-霍伊尔（Bondi-Gold-Hoyle）稳恒态宇宙理论。真是太好了！罗杰想，也许席艾玛能解决自己的疑惑。“霍伊尔说，照稳恒态理论，宇宙的膨胀将把一些遥远星系赶出我们的视线；星系将运动到我们看不见的地方。但我不明白这是如何发生的。”罗杰拿出笔来开始在餐巾纸上画时空图，“根据这个图，我想星系会越来越暗，越来越红，但永远不会完全消失。我哪儿错了吗？”

　　席艾玛大吃一惊，他从没见过一幅图能有那么大的威力。彭罗斯是对的，霍伊尔一定错了。更重要的是，奥利弗的弟弟真了不起。

　　于是，席艾玛和罗杰开始研究这种图，后来在60年代，他还将和他自己的学生（霍金、埃里斯、卡特尔、里斯等等，见第7章）继续研究。他拉着彭罗斯讨论了好几个小时，谈一些发生在物理学中激动人心的事情。正在发生的事情，席艾玛都知道，他自己的热情和兴奋也感染了彭罗斯。很快，彭罗斯就被钩住了。他要完成他的数学博士，不过对宇宙的追求从此也成为他向前的动力。在未来的10年里，他一只脚牢牢地扎根在数学，另一只脚则踏进了物理学。



罗杰·彭罗斯，约1964年。[Godfrey Argent为英国国家肖像画廊和伦敦皇家学会摄。Argent提供。]

新思想常出现在一些奇怪的时刻，出现在人们最意想不到的时候。我想，这是因为新思想来自人的潜意识，在意识不太活跃的时

候，潜意识最有力量。一个好的例子是，霍金1970年在正准备睡觉时，发现了黑洞视界的面积必然总是增大的（第12章）。另一个例子就是彭罗斯改变我们对黑洞内部认识的发现。

1964年晚秋的一天，[15]伦敦伯克贝克学院教授彭罗斯和朋友罗宾逊（Ivor Robinson）正向办公室走去。在过去的一年里，类星体发现了，天文学家开始猜测它们的能源来自星体坍缩（第9章）。自那时起，彭罗斯也试图弄清楚，奇点是否是由真实的随机形变的坍缩恒星产生的。当他和罗宾逊边走边谈时，潜意识正在思考令他疑惑的那些问题——他的思想已跟它们搏斗好久好久了。

彭罗斯回忆说："过一条马路时，我和罗宾逊停止了说话。过去以后，又接着谈。显然，在过马路那会儿，我忽然想到了什么东西，但后来的谈话又把它从脑子里赶走了！那天晚些时候，罗宾逊走了，我回到办公室。我记得有过一种奇怪的欣喜，但说不清那是什么。我开始在脑子里搜寻那天想过的事情，想找出令我欣喜的那样东西。排出许许多多不太可能的事情后，我终于发现了过街时产生的思想。"[16]

这思想很美，和以前在相对论物理学中见过的任何事物都不一样。在接下来的几个星期里，彭罗斯认真清理了他的思想，从不同方向去审视它，完成它的细节，尽可能使它具体，在数学上精确。一切理顺后，他为《物理学评论通讯》写了一篇短文，[17]描述了星体坍缩中的奇点问题，然后证明了一个数学定理。

彭罗斯的定理大意说，假定一颗恒星——不论什么样的恒星——发生了剧烈坍缩，使引力变得足够强大从而形成一个显视界，就是说，强大的引力足以将外出的光线拉回来（卡片12.1）。显视界形成以后，不可能有什么东西能阻挡引力进一步增强而产生奇点。结果（由于黑洞总有显视界），每个黑洞在它内部都必然有一个奇点。

这个奇点定理最令人惊讶的地方在于它巨大的普适能力。它不仅

适用于具有特殊理想化性质的（如完全球状的或没有压力的）坍缩恒星，也不仅适用于初始随机形变很小的恒星，而且适用于一切可以想象的坍缩恒星，这样，它无疑也适用于我们现实宇宙中存在的真实的坍缩恒星。

彭罗斯奇点定理的惊人力量来自他在证明中使用的一种新的数学工具，以前还没有物理学家在弯曲时空的计算（也就是广义相对论的计算）中用过，那就是拓扑学。

拓扑学是研究事物间或事物自身的定性联系方式的数学分支。举例说，咖啡杯和炸圈饼“有相同的拓扑”，因为（假如它们都是用泥做的）我们可以光滑而连续地将一个变形为另一个，而不会撕裂它，即不会改变任何联系（图13.5（a））。相反，球与炸圈饼有不同的拓扑，为把球变成饼，我们必须在球上挖一个洞，从而改变了它自己的联结方式（图13.5（b））。

拓扑学只关心联结，不关心形状、大小和曲率。例如，炸圈饼和咖啡杯的形状和曲率是大不相同的，但它们有相同的拓扑。[2]

我们物理学家在彭罗斯奇点定理之前忽略了拓扑学，因为我们固执地认为，时空曲率是广义相对论的中心概念，而拓扑学不能告诉我们任何有关曲率的事情。（实际上，因为彭罗斯的定理对拓扑学的依赖太强，它没有为我们带来关于奇点曲率的东西，就是说，没有关于奇点的潮汐引力的细节的东西。定理只告诉我们，在黑洞内的某个地方，时空到达了终点，到达那个终点的任何事物都会被破坏。如何破坏是曲率的事；它们必然遭到破坏，时空总会遇到终点——这个事实是拓扑学关心的。）

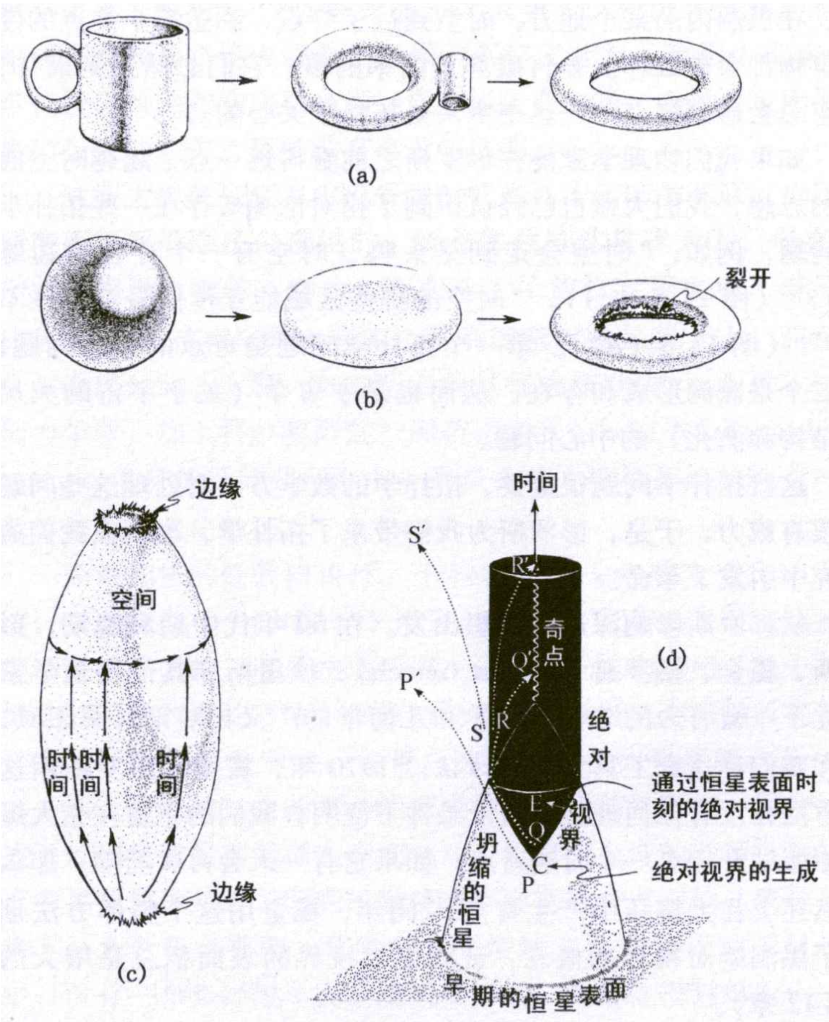



图13.5　下面的所有问题都是关于点的联络性质的，即它们是拓扑学的问题。（a）咖啡杯（左）和炸圈饼（右）能光滑连续地互相变形为对方而不会出现破裂，就是说，不会改变任意两点联系的定性特征。这样，它们具有相同的拓扑。（b）为把球（左）变形为饼（右），必须在球上打一个洞。（c）这里画的时空有两个尖锐边缘（类似于（b）的裂口），一个是时间的起点（我们宇宙大爆炸的开端），另一个是时间的终点（类似于大挤压）。我们也可以想象一个在所有时间里存在而且总会继续存在的宇宙，这样宇宙的时空没有边缘。（d）黑色的空间区域是黑洞的内部；白色区域是黑洞的外部（见卡片12.1）。内部的点不可能向外面的点传送信号。

如果我们物理学家能在彭罗斯之前看得远一点，超越时空曲率的

思想，我们大概也已经认识到了相对论确实存在一些拓扑学的问题，例如，“时空会走到尽头吗（时空有一个存在的边缘吗）？”（图13.5（c））,“时空的哪些区域能互传信号？哪些不能？”（图13.5（d））。第一个拓扑学问题是奇点的中心问题；第二个是黑洞形成和存在，从而也是宇宙学（关于宇宙的大尺度结构和演化）的中心问题。

这些拓扑学问题很重要，拓扑学的数学方法对处理这些问题也很有威力，于是，彭罗斯为我们带来了拓扑学，也就在我们的研究中引发了革命。

从彭罗斯影响深远的思想出发，在60年代中期和晚期，彭罗斯、霍金、格罗赫（Robert Geroch）、埃里斯和其他物理学家创立了一套有力的结合拓扑学和几何学的广义相对论计算工具，现在我们称这套工具为整体方法。[18]1970年，霍金和彭罗斯用这种方法在没有任何理想化假设条件下证明，我们的宇宙在它大爆炸膨胀的开端有一个时空奇点；如果它有一天会再次坍缩，那么必然还会在大挤压中产生奇点。[19]同年，霍金用这个整体方法创造了黑洞绝对视界的概念，证明绝对视界的表面积总是增大的（第12章）。

现在，让我们转到1965年，来看一场重要的论战。卡拉特尼科夫和栗弗席兹在莫斯科证明（他们是这样想的），在内部随机变形的扰动下，真实恒星坍缩成黑洞时不可能在黑洞中心产生奇点；而彭罗斯在英国证明，每个黑洞在中心都必然有奇点。

演讲厅有250个座位，卡拉特尼科夫站起来讲话时，人已经坐不下了。那是在1965年夏的一个大热天，全世界重要的相对论研究者聚集在伦敦召开第三届国际广义相对论和引力论会议。卡拉特尼科夫和栗弗席兹第一次有机会在这样一个世界范围的集会中报告他们关于黑洞没有奇点的证明。

从斯大林死后到戈尔巴乔夫时代的几十年间的苏联，出国申请能否得到批准是很难说的。栗弗席兹虽然是犹太人，但在50年代是相当自由的，然而，现在他上了旅行者黑名单，得等到1976年。[20]卡拉特尼科夫有两个不利因素，他是犹太人，而且从来没有出过国。（第一次申请出国是很难获准的。）不过，经过努力争取，加上科学院副院长谢苗诺夫（Nikolai Nikolaievich Semenov）为他给苏联共产党中央委员会打了电话，卡拉特尼科夫总算被批准来伦敦了。[21]

面对拥挤的伦敦演讲厅，卡拉特尼科夫拿着麦克风，一边讲，一边在占了一面墙的15米宽的黑板上写满了方程。他的方法不是拓扑学的，而是物理学家在分析时空曲率时用了几十年的标准的满是方程的方法。卡拉特尼科夫从数学上说明随机扰动一定会随恒星坍缩而增大。他说，这意味着，如果坍缩会形成奇点，那么奇点的时空曲率必然遭受了彻底的随机形变。接着他讲述了他和栗弗席兹怎样在广义相对论定律所允许的各种类型的奇点中寻找那种经过了彻底的随机曲率形变的奇点。他从数学上例举了一个个奇点类型，几乎令人厌烦地为这些奇点编了目。其中，没有一个是经过了完全随机变形的。于是他得到结论——也结束了他40分钟的讲话——随机扰动下的坍缩恒星不可能产生奇点。扰动一定能将恒星从毁灭中救出来。

掌声响过，惠勒最有才能的学生米斯纳站起来，奋力提出反驳。他很激动，又精力旺盛，用连珠炮似的英语讲述了彭罗斯几个月前证明的定理。如果彭罗斯的定理是对的，那么卡拉特尼科夫和栗弗席兹就一定错了。

这位苏联代表愣了，激动了。米斯纳的英语说得太快，他没跟上。又由于彭罗斯的定理靠的是相对论专家们都很陌生的拓扑学论证，所以苏联人认为那是可疑的。相反，卡拉特尼科夫-栗弗席兹分析的基础很牢固，他们断言，彭罗斯可能错了。[22]

接下来的几年里，东西方的相对论专家彻底考察了彭罗斯和卡拉

特尼科夫-栗弗席兹的分析，乍看起来，两家都可疑，都有可怕的潜在裂痕。不过，随着专家们逐步掌握和发展彭罗斯的拓扑学技术，他们相信彭罗斯是对的。



1971年6月，在莫斯科卡拉特尼科夫家里的晚餐聚会。左起顺时针方向：索恩、惠勒、栗弗席兹、卡拉特尼科夫、卡拉特尼科夫的夫人Valentina Nikolaievna ，别林斯基和卡拉特尼科夫的女儿Eleanora。[C·W·米斯纳提供。]

1969年9月，我在莫斯科泽尔多维奇研究小组访问，栗弗席兹给我拿来一份他和卡拉特尼科夫刚写好的手稿。“基普，请为我把这份稿子带到美国，交给《物理学评论通讯》。”[23]他解释说，在苏联写的稿子，不论什么内容，都自动划为机密，等解密以后才能拿出来，而那得等三个月。照苏联可笑的体制，我们这些外国访问者可以在莫斯科阅读这些手稿，但不经审查不得将稿子带出国。这篇稿子太可贵了，需要尽快发表，不能让那种荒唐的审查给耽误了。栗弗席兹告诉我，他们在文章里认输了，为错误而遗憾：彭罗斯是对的，他们错了。1961年，他们没能在爱因斯坦场方程的解中找到任何带有完全随机形变的奇点；但现在，受彭罗斯定理激发，他们和研究生别林斯基（Vladimir Belinsky）设法找到了一个。他们认为，这个新奇点可能终结随机形变恒星的坍缩，也可能最终在大挤压的终点毁灭我们的宇

宙。实际上，我在1993年想，他们可能是对的。在本章临近结束的时候，我还会来谈1993年的观点和他们新的RKL（“Belinsky-Khalatnikov-Lifshitz”）奇点的性质。

我自己有过体会，对一个理论物理学家来说，承认发表的结果犯存重大错误，不仅仅是难堪的事情，那差不多是自我毁灭了。1966年，我的白矮星脉动计算错了，两年后，我的错误计算害得天文学家们把新发现的脉冲星当成白矮星。错误发现以后，因为意义重大，所以在英国的《自然》杂志的编者按里特别指出来。那真是一粒难咽的药丸。

像这样的错误，在美国和欧洲能损害一个物理学家，在苏联就严重得多了。在苏联，科学家在这一群体中的社会地位是特别重要的，它关系着能否出国，关系着科学院的院士选举，这些又可能带来特权，如比别人高近一倍的工资，专用小轿车等。在这样的一些诱惑下，苏联科学家比西方的科学家更容易隐瞒和逃避错误。所以我才被栗弗席兹的请求所感动。他不愿意真理的传播受到阻碍，他的手稿也写得很坦诚：承认错了，并宣布未来的一版《经典场论》（朗道-栗弗席兹的广义相对论教程）将修正黑洞不产生奇点的结果。

我把手稿藏在我个人的论文中间带回美国，把它发表了。[24]苏联的当权者们从未发觉。

为什么把拓扑学方法带进相对论研究的是一个英国物理学家（彭罗斯），而不是美国、法国或苏联的哪个物理学家？为什么整个60年代中拓扑学方法在英国相对论物理学家中间如火如荼，而在美国、法国、苏联和其他地方却举步维艰呢？

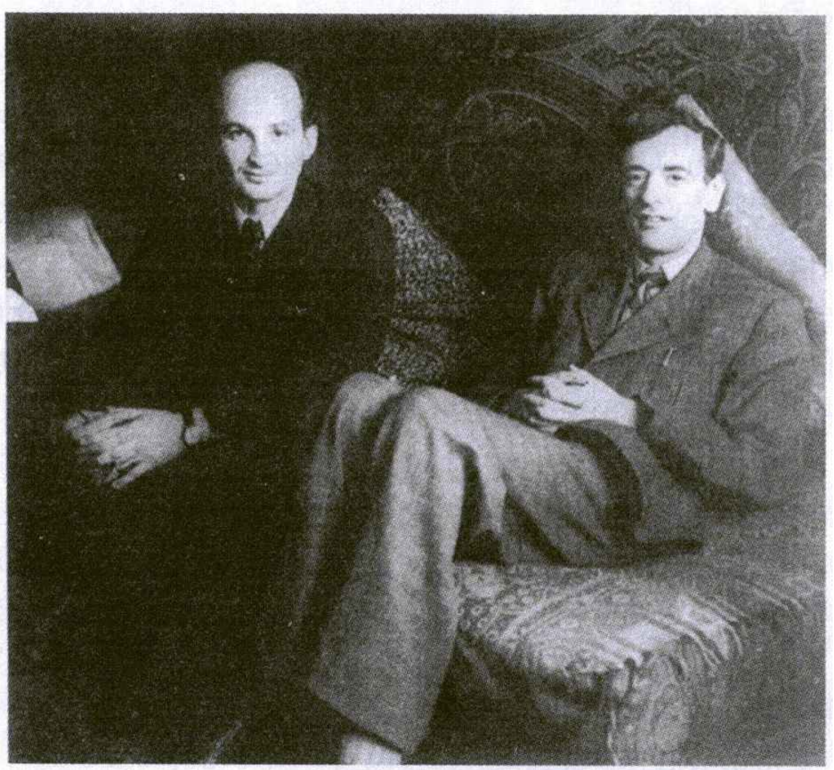
我想，原因在于英国理论物理学家在大学所受的教育。他们在大学时主要学数学，然后在应用数学系或应用数学和理论物理学系做博士研究。在美国却不同，大物理学家们在大学时一般都以学物理为

主，然后在物理系做博上研究。这样，年轻的英国理论物理学家会很熟悉那些还没在物理学中应用过的艰深的数学分支，似对像关于分子、原子和原子核行为的那些“大胆的”课题，他们也可能缺乏良好的基础。反过来，年轻的美国物理学家在数学上不比他们的物理学教授知道页多，但在分子、原子和原子核的特殊问题上，他们是游刃有余的。

二战以来，美国在很大程度上成了理论物理学的主角，但我们也向全世界的物理学同行们暴露出令人惭愧的数学水平。我们多数还在用50年前的数学；没有能力和现代数学家交流。由于没受过多少数学训练，我们美国人很难在彭罗斯引进拓扑学方法时学会运用它们。

法国物理学家所受的数学教育甚至比英国还好。然而，在六七十年代，法国的相对论专家们被数学的严格（也就是完美）所纠缠，不太重视物理直觉，所以他们没能为坍缩恒星和黑洞的认识做多少事情。对严格数学的追求拖住了他们向前的脚步，尽管他们很熟悉拓扑学，但也无法同英国人竞争。他们甚至没有一点儿这方面的尝试，精力都被吸引到别的地方去了。

从30年代到60年代，朗道在很大程度上代表着苏联的理论物理学，他也是苏联抵制拓扑学的主要根源：30年代，他把理论物理学从西欧带回苏联（第5章）。他的一个传播工具是他创立的一套理论物理学考试，叫“理论须知”，想进他的研究小组的人都得通过这个考试。任何人，不论原来学什么，都可以来参加考试，但很少有人通过。在“理论须知”实行的29年（1933~1962）里，只有43人过了，他们的很大一部分后来都有重大的物理学发现。[25]



栗弗席兹（左）和朗道（右），1954年在莫斯科绍瑟街2号物理问题研究所的朗道家里。[栗弗席兹夫人Zillaida Ivanorna提供。]

“理论须知”中的数学问题来自所有朗道认为对理论物理学重要的数学分支，覆盖了微积分、复变函数、微分方程定性理论、群论和微分几何，都是物理学家一生所需要的。其中没有拓扑学，不是朗道反对它，而是忽略了它，认为它没有关系，用不着它——他的观点也就几乎成了40年代到60年代大多数理论物理学家信奉的真理：拓扑学与理论物理毫不相干。

这种观点，通过朗道和栗弗席兹写的一套《理论物理学教程》传给了全世界的理论物理学家。那是20世纪在世界范围内最有影响的物理学教科书，不过也跟朗道的理论须知考试一样，忽略了拓扑学[3]。

奇怪的是，早在彭罗斯定理之前，列宁格勒的两个苏联数学家亚历山大洛夫（Aleksander Danilovich Aleksandrov）和皮苗诺夫（Revol't Ivanovich Pimenov），就在相对论研究中应用拓扑技术了。[26]1950～1959年，亚历山大洛夫用拓扑学探求时空的“因果结构”，也就是研究能相互通讯和不能相互通讯的时空区域之间的关系，[27]这就是后来在黑洞理论中大获丰收的那种拓扑分析方法。他建立了一个优美而有力的拓扑学框架，50年代中期，他的年轻同事皮苗诺夫又接着把它向前推进。[28]

但他们的研究没有结果。在与亚历山大洛夫和皮苗诺夫往来的物理学家中，几乎没有引力方面的专家。本来，这样的专家会知道哪些计算有用，哪些没有用，他们会告诉亚历山大洛夫和皮苗诺夫，大爆炸奇点和恒星的引力坍缩正需要用他们的方法去探索。但是，在列宁格勒听不到这样的忠告；他们需要的物理学家远在东南600公里的莫斯科，而那些人从没想过拓扑学和拓扑学家。亚历山大洛夫-皮苗诺夫拓扑结构，刚开花就凋落了。

花的命运是和两个主人的命运联系在一起的。亚历山大洛夫成为列宁格勒大学校长，没有更多的时间做进一步研究；皮苗诺夫因为建立“反苏组织”在1957年被捕，坐了6年牢，出来7年后又被捕了，流放5年，去了列宁格勒以东1200公里外的科密共和国。

我没见过亚历山大洛夫和皮苗诺夫，但我1971年（皮苗诺夫第二次被捕后一年）访问列宁格勒时，皮苗诺夫的故事还在物理学家中流传。据说，皮苗诺夫认为苏联政府道德败坏，像美国许多年轻人在越南战争时期的思想一样，他觉得与政府合作就是同流合污。保持自己道德纯洁的惟一办法是非暴力对抗。在美国，非暴力对抗意味着拒绝登记当兵；皮苗诺夫的非暴力对抗则是“萨密兹达”（Samizdat），也就是“地下出版”禁止的手稿。有人说，皮苗诺夫常从朋友那儿收到禁止在苏联出版的一些稿子，他用复写纸抄几份，然后把它们寄给别的朋友，他们也如法炮制，继续扩散。皮苗诺夫被捕了，被判有罪，流

放到科密，在那儿做伐木工人和伐木机厂的电机技师。后来，科密科学院发现了他，让他做数学部主任。

终于又能做数学了，皮苗诺夫继续研究他的时空拓扑学。那个时候，拓扑学作为物理学家的引力研究工具已经生根了，而他却孤独地远离他祖国的前沿物理学家。他没有产生影响；假如换个环境，他是应该能够产生影响的。

与亚历山大洛夫和皮苗诺夫相比，彭罗斯要幸运多了。他一只脚牢牢扎根在数学，另一只脚牢牢扎根在物理，这是他成功的重要原因。

[1] 1828年，伦敦大学建立时，也叫伦敦大学学院；1898年，大学学院成为伦敦大学众多学院的一个。——译者

[2] 拓扑或拓扑学（topology）是音译名词；英语词源来自希腊语的tópos，意思是“位置”，的确反映了这个学科的特征。——译者

[3] 朗道的“理论须知”（后来叫“朗道位垒”）包括，数学：解常微分方程（任选一题），用初等函数表示不定积分（任选一题）；物理：理论力学，热力学和统计物理，场论（狭义和广义相对论），非相对论量子力学，相对论量子力学、场论、基本粒子，连续介质电动力学，连续介质力学（流体力学、弹性力学）。必须的数学都包括在物理学题目中。通过了这个“位垒”的人也有永远离开物理学的，可见那是多么痛苦的考试。由朗道设计，主要由栗弗席兹执笔的那套教程计划八卷（现在我们看到的有10卷），即须知的那些物理内容，几乎都有中译本（但多数都是根据早期版本翻译的）。——译者

## 最佳猜想

也许有人以为，彭罗斯的奇点定理将一劳永逸地解决黑洞内部的问题。事情没那么简单；它反倒引出了一些新问题——从60年代中期以来，物理学家一直在同这些问题斗争，但成绩不大。关于这些问题，我们现在（1993年）的回答（更好的说法是，我们的“最佳猜想”）是：

1.进入黑洞的一切事物都必然会被奇点吞没吗？我们认为是的，但还不能肯定。

2.存在从黑洞内部到其他宇宙或我们宇宙的其他部分的道路吗？很可能没有，但我们没有绝对把握。

3.落进奇点的事物的命运是什么？我们认为，在黑洞很年轻时下落的事物在量子引力发生重要作用之前会被潮汐引力任意猛烈地撕裂。然而，落向老黑洞的事物可能会幸免于难，而最终能够面对量子引力定律。

在这一章剩下的篇幅里，我将更详细地解释这些答案。

回想一下，对上面三个问题，奥本海默和斯尼德曾有过明确的回答：如果黑洞是由高度理想化的球状坍缩恒星产生的，那么，（1）进入黑洞的一切事物都会被奇点吞没；（2）没有到其他宇宙或我们宇宙的其他部分去的事物；（3）接近奇点时，一切事物都将遭受无限增大的径向拉伸和横向挤压（图13.1上）从而被毁灭。

这些答案是很有启发作用的，它激发了能带来更深刻认识的计算。不过，（科拉特尼科夫和栗弗席兹得到的）更深刻的认识却表明，奥本海默-斯尼德答案与我们生存的真实宇宙无关，因为发生在所

有真实恒星的随机变形会彻底改变黑洞的内部。奥本海默-斯尼德黑洞的内部“对小扰动是不稳定的”。[29]

爱因斯坦场方程的雷斯纳-诺德斯特勒姆解也提出了明白而确凿的答案：假如黑洞是由特殊的高度理想化的球状带电恒星产生的，那么，坍缩的恒星和落进黑洞的其他事物可以经过一个“封闭小宇宙”从黑洞内部旅行到另一个大宇宙去（图13.4）。[30]

这个答案也有启发意义（它已经为科幻小说家们提供了好多想象的素材）。然而，它跟奥本海默-斯尼德预言一样，与我们生存的真实宇宙无关，因为它对小扰动是不稳定的。说得具体一些，在我们的现实宇宙中，黑洞不断遭受微弱电磁真空涨落和少量辐射的轰击。这些涨落和辐射落进黑洞，在黑洞引力作用下加速，增大能量，然后猛烈地打击并摧毁还没来得及启程的封闭小宇宙。这个猜想是彭罗斯1968年提出的，自那时起，得到了许多物理学家不同计算的证明。[31]

另外，别林斯基、卡拉特尼科夫和栗弗席兹还为我们的问题提出了一个答案，它对小扰动是完全稳定的，也许就是适用于我们宇宙中真实黑洞的“正确”答案：形成黑洞的恒星和在黑洞年轻时落进洞里的一切事物都会被BKL奇点的潮汐引力撕裂。（这是别林斯基、卡拉特尼科夫和栗弗席兹相信彭罗斯说的黑洞内一定存在奇点后，从爱因斯坦方程的解中发现的一类奇点。）[32]

BKL奇点的潮汐引力与奥本海默-斯尼德奇点根本不同。奥本海默-斯尼德奇点越来越强地作用在下落的宇航员（或其他任何事物）上，总是径向拉伸，横向挤压，而且拉伸和挤压的强度持续光滑地增大（图13.1）。BKL奇点不一样，它有点像我们在糖果店和博览会上见过的制糖机，先在一个方向挤压，然后又在另一个方向挤压；一会这儿，一会那儿，（在下落宇航员看来），拉伸与挤压总在随机地不可捉摸地变换方向，平均说来，作用强度也越来越大，当宇航员离奇点越来越近时，他会感觉振荡也越来越快。米斯纳（他独立于别林斯

基、卡拉特尼科夫和栗弗席兹，也发现了这种混沌振荡的奇点）称这种行为是搅拌振荡。[33]我们可以想象，它像打蛋器搅拌蛋黄蛋白那样将宇航员身体的各部分搅在一起。图13.6表现了潮汐引力如何振荡的一个具体例子，但精确的振荡序列是混沌而难以预料的。

米斯纳的搅拌型奇点的振荡在（从宇航员看来）某个特别时刻在空间各处都是一样的。BKL奇点却不像这样，它的振荡在空间和时间上都是混沌的，就像破碎的海浪的前锋在时空中的湍流运动。例如，当宇航员的头在南-北方向遭受轮番拉伸和挤压时，它的右脚可能正在东北-西南方向受罪，而左脚却在南-东南和北-西北方向遭殃。他的头、左脚和右脚“挨打”的频率可能会大不相同。

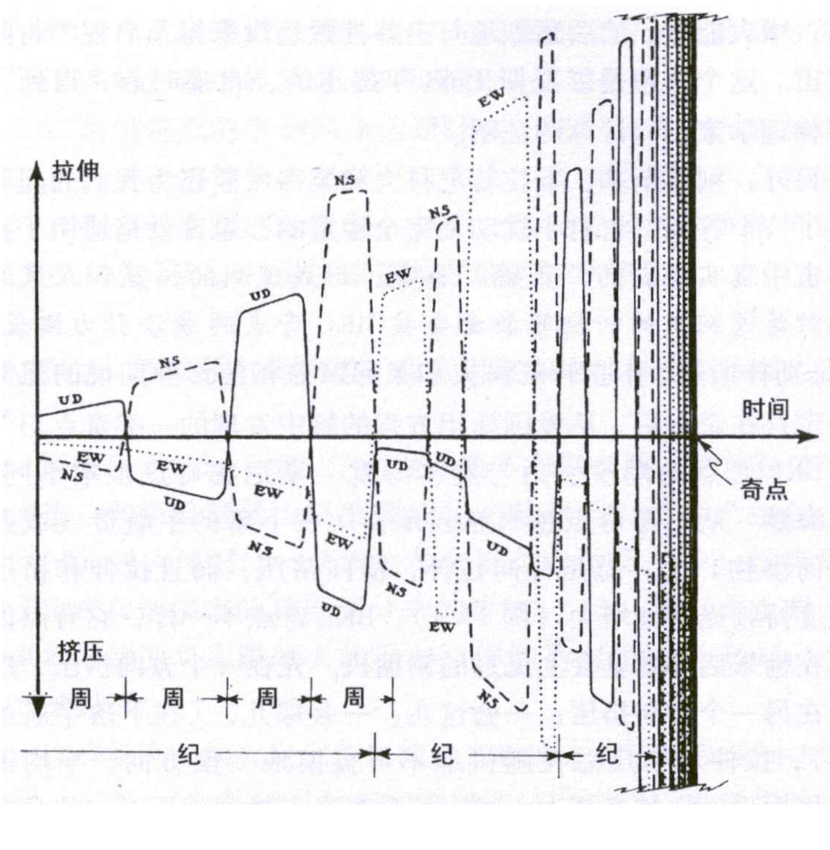


图13.6　BKL奇点的潮汐引力随时间振荡的例子。潮汐力以不同方式作用在三个互相垂直的方向。为明确起见，我们称三个方向为UD（上-下）、NS（南-北）和EW（东-西），三条曲线分别描述三个方向的潮汐引力作用。时间为水平方向。UD曲线在水平时间轴上方时，潮汐力沿UD方向拉伸，UD曲线在时间轴下方时，潮汐力产生挤压。曲线在轴上越高，拉伸作用越强；在轴下越低，挤压作用越强。注意以下几点：（i）在任何时刻，挤压发生在两个方向，而拉伸在一个方向。（ii）潮汐力在拉伸与挤压间振荡，每次振荡叫“周”。（iii）周合成“纪”。在每一纪里，一个方向只有完全持续的挤压，另外两个方向在拉伸与挤压间振荡。（iv）纪改变时，持续挤压方向也发生改变。（v）临近奇点时，振荡无限快，潮汐力无限大。周划归为纪的细节和振荡模式在每一纪开始时的改变，可以用所谓的“混沌图”来描述。

爱因斯坦方程预言，宇航员到达奇点时，潮汐力长到无限大，混沌振荡变得无限快。宇航员死了，构成他身体的原子遭到了无限的破坏，混沌地搅在一起——这时，一切事物（潮汐力、振荡频率、破坏、混合）都成为无限，时空也不复存在。

然而，量子力学定律不同意。它们严禁无限。就我们现在（1993年）的认识，在邻近奇点处，量子力学的定律和爱因斯坦广义相对论的定律会融合在一起，将彻底改变“游戏规则”。新定律叫量子引力。

量子引力发生作用的时候，宇航员死了，他身体的各部分完全混合了，原子被彻底破坏，不能识别了。但没有一样是无限的，“游戏”还可以继续。

那么，量子引力到底什么时候发生作用，它能做什么呢？据我们现在的认识（很可怜的一点认识），当振荡的潮汐引力（时空曲率）大得能在$10^{-43}$秒或更短时间内彻底改变所有物体的时候，[1]量子引力就出现了。[34]接着，它将根本改变时空的特性：它分裂空间和时间统一而成的时空；它分开胶结在一起的空间和时间；它毁灭时间的概念，也破坏空间的确定性。时间不在了，我们不能再谈什么“这件事情发生在那件事情以前”，因为没有时间，就没有“以前”和“以后”的概念。统一时空惟一遗留下来的空间，成了像肥皂泡一样随机的概率的泡沫。[36]

分裂前（也就是在奇点外）的时空，可以比喻为一块饱含水的木

头，木块是空间，水是时间。两样东西（木头和水；空间和时间）是紧紧交织在一起的、统一的。时空走近奇点和量子引力，就像木块投进烈火。火把木块里的水蒸发出去，只留下脆弱的干木块；在奇点，量子引力毁灭了时间，只留下脆弱的空间。接着，木块燃烧了，成为一堆烟灰；而量子引力则把空间变成一团随机的概率的泡沫。

这些随机的概率的泡沫就是在量子引力定律作用下构成奇点的东西。泡沫的空间没有任何确定的形状（也就是没有确定的曲率，甚至没有一定的拓扑），它只有这样那样形状（即这样那样曲率和拓扑）的概率。例如，在奇点内，空间具有如图13.7（a）那样的曲率和拓扑的概率可能是0.1%，图13.7（b）的概率是0.4%，图13.7（c）的概率是0.02%，等等。这并不是说，空间用它时间的0.1%处于形状（a），用时间的0.4%处在（b），用时间的0.02%处在（c），因为在奇点内没有时间这种东西。同样，因为没有时间，像空间形状（b）处于（c）“以前”还是“以后”那样的问题也完全是没有意义的。关于奇点，我们能提出的惟一有意义的问题只能是，“构成你的空间在形状（a）、（b）和（c）的概率有多大？”答案很简单，0.1%，0.4%和0.02%。

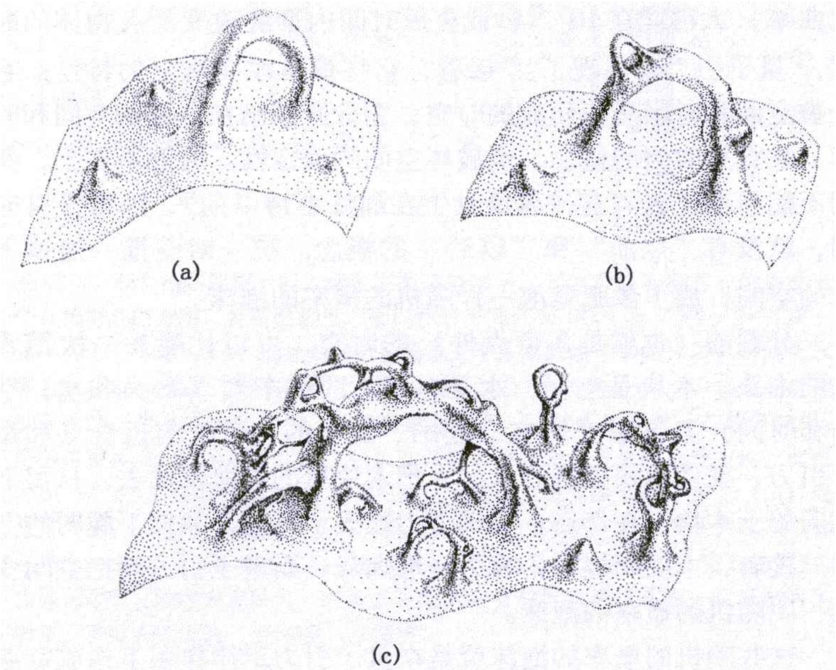



图13.7　我们猜想存在于黑洞内奇点的量子泡沫的嵌入示意图。空间的几何和拓扑是不确定的，而是一些概率。例如，图中（a）所示的空间形状的概率是0.1%，（b）的概率是0.4%，（c）的概率是0.02%，等等。

在奇点内，任何可以想象的曲率和拓扑都允许存在，所以，尽管听起来有些荒唐，我们还是说奇点由概率泡沫构成，惠勒称它为量子泡沫，[37]他第一个提出，在量子引力定律作用下，空间性质一定是这样的。

概括地说，在黑洞中心，在BKL潮汐引力振荡到达极点的时空区域，存在着一个奇点：在那儿，时间不复存在，空间成了量子泡沫。

量子引力定律的一个使命是决定黑洞奇点内不同曲率和拓扑的概率。另一个使命也许是决定奇点诞生“新宇宙”的概率，也就是奇点像大爆炸奇点在约150亿年前产生我们的宇宙那样，产生一个新的经典的（非量子的）时空区域。

黑洞奇点产生“新宇宙”，可能吗？我们不知道。也许，它永远不会发生；也许，它经常发生——也许，我们相信奇点是由量子泡沫构成的，可能是完全跟错了方向。

现在，霍金、哈特尔和其他一些人正在惠勒和德维特奠定的基础上进行研究，2在未来的10年或20年里，他们也许能给出一个明确的回答。[38]

宇宙万物都会老：恒星燃尽燃料而死亡；地球最终失去大气而成为死星；我们人类也会满脸皱纹而更老练。

黑洞深处奇点附近的潮汐引力也不例外。根据艾伯塔大学伊斯雷尔和泊松（Eric Poisson）以及在加州理工学院小组（在以前多罗什科维奇（Andrei Doroshkevich）和诺维科夫工作基础上建立的）的博士后奥里（Amos Ori）1991年的计算，它们也是随年龄而改变的。黑洞初生时，内部潮汐力表现出剧烈混乱的BKL式振荡（图13.6上）。然而，随着黑洞变老，随机振荡也更平稳柔和，并逐渐消失。[39]

例如，某个类星体中心有一个100亿个太阳质量的黑洞，在它诞生几个小时后落进来的宇航员会被疯狂振荡的BKL潮汐引力撕裂。然而，等了一两天才落进来的第二个宇航员所遭遇的振荡的潮汐力就温和得多。当然，潮汐力的拉伸和挤压作用还是足以杀死他，不过比一天前的痛苦轻得多，他可以多活些时候，可以比第一个宇航员走得离奇点更近一些。第三个宇航员等了几年才进来，他的遭遇更加温和。照伊斯雷尔、泊松和奥里的计算，奇点周围的潮汐力，这时已经相当轻柔了，宇航员几乎感觉不到。他会活下来，也许还不受一点儿伤害，他能走到随机量子引力奇点的边缘。不过当他在奇点边缘直面量子引力定律时，还是会被杀死——我们也没有绝对把握说他一定会在那儿死，因为我们还根本没有很好地认识量子引力定律和它们的结果。

黑洞内潮汐力的衰减并不是无法改变的。任何时候落进黑洞的物质和辐射（或宇航员）总会给潮汐力增添能量，这些东西就像一块扔给狮子的肉。奇点附近振荡的拉伸和挤压作用得到补给后，会在短时间内增强，然后又衰减下去，回归刚才的宁静。

惠勒在50年代末和60年代初有一个梦想，一个希望：人类有一天能走进奇点去看量子引力如何发生作用——这样，我们不仅能靠数学和计算机模拟来研究它，还可以凭借真实的物理观测和实验。奥本海默和斯尼德令这个梦想破灭了（第6章）。他们发现，坍缩恒星周围形成的视界把奇点藏了起来，不让外面看到，假如我们总在视界外面，就没有办法探索奇点。假如我们穿过一个巨大的老黑洞的视界，活着面对量子引力奇点，我们也没有办法把看到的情况传回地球。我们的信息逃不出黑洞；视界把它遮住了。

虽然惠勒早就不做那样的梦了，现在也热情主张不可能走近奇点，但这一点是否正确，我们还完全没有把握。可以想象，某些极端的非球状星体坍缩会产生裸奇点，即没有视界包围的奇点，从而可以从外面的宇宙，甚至从我们的地球观察它、探索它。

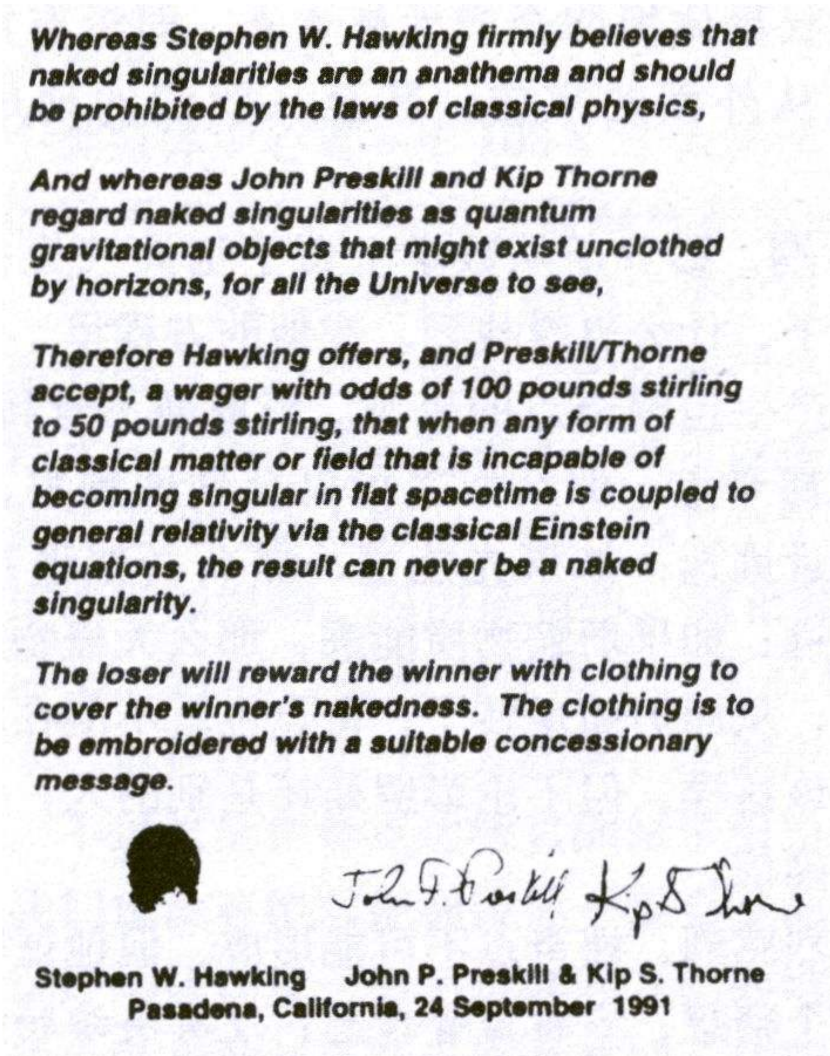
60年代后期，彭罗斯从数学上费了很大气力去寻找产生裸奇点的坍缩例子，什么也没找到。在他的方程里，每当坍缩产生奇点，它总会产生包围奇点的黑洞。彭罗斯不觉得奇怪。毕竟，假如真会形成裸奇点，那么似乎可以合理地预料，在奇点形成前，光能从附近逃逸；如果光能逃，那么（似乎）产生奇点的坍缩物质也能逃；如果坍缩物质能逃，那么大概物质内部的巨大压力会让它逃，从而坍缩逆转了，奇点也就不能在原来的地方形成了。似乎应该这样，但不论彭罗斯还是别的人，他们的数学还没有能力让人确信。

彭罗斯强烈感到，裸奇点不可能形成，但他证明不了。1969年，他提出一个猜想，宇宙监督猜想：没有坍缩物体能形成裸奇点；如果奇点形成了，它必然套在视界里，我们不能从外面的宇宙看见它。

物理学的“建设者”们——像惠勒那样的物理学家，他们的观点总是最有影响的——接受了宇宙监督，几乎把它当成真理了。不过，自彭罗斯提出四分之一世纪以来，宇宙监督还没得到证明；而最近的计算机对高度非球状星体坍缩的模拟甚至令人怀疑它可能是错的。根据康奈尔大学夏皮罗和特奥科尔斯基的这些模拟，有些坍缩确实可能会产生裸奇点。[40]可能产生而不是一定会产生，不过可能而已。

霍金是当今物理学建设者的缩影，而普雷斯基尔（John Preskill，我在加州的同事）和我喜欢给他们的建设加把劲儿。于是，我们在1991年跟霍金打赌（图13.8），我们赌宇宙监督是错的，裸奇点能在宇宙中形成；霍金赌它是对的，裸奇点永远不会形成。

赌约才立四个月，霍金自己就发现数学证据（但不是严格的证明）表明，黑洞在完全蒸发后（第12章），可能不会像他以前预料的那样彻底消失，而会留下一个小的裸露的奇点。[41]几天后，在普雷斯基尔家聚餐时，他把结果告诉了普雷斯基尔和我。不过，当我俩要他认输时，他却不肯，找了一个技术上的根据。他说，赌约写得很清楚，我们的赌限于在包括广义相对论的经典物理学（也就是非量子的）定律作用下形成的裸奇点。但是，黑洞蒸发是量子力学现象，不受经典的广义相对论定律作用，而是由弯曲时空的量子场论定律决定的，所以任何可能从黑洞蒸发产生的裸奇点都在我们的赌约范围之外。霍金说对了。不管怎么说，无论裸奇点如何产生，它总是对物理学建设的一个打击！


Whereas Stephen W. Hawking firmly believes that naked singularities are an anathema and should be prohibited by the laws of classical physics,

And whereas John Preskill and Kip Thorne regard naked singularities as quantum gravitational objects that might exist unclothed by horizons, for all the Universe to see,

Therefore Hawking offers, and Preskill/Thorne accept, a wager with odds of 100 pounds stirling to 50 pounds stirling, that when any form of classical matter or field that is incapable of becoming singular in flat spacetime is coupled to general relativity via the classical Einstein equations, the result can never be a naked singularity.

The loser will reward the winner with clothing to cover the winner's nakedness. The clothing is to be embroidered with a suitable concessionary message.

Stephen W. Hawking     John P. Preskill & Kip S. Thorne

Pasadena, California, 24 September 1991

图13.8　霍金、普雷斯基尔和我为彭罗斯的宇宙监督猜想的赌约。3

虽然打赌好玩儿，我们讨论的东西却是很严肃的。如果裸奇点能够存在，那么只有在我们现在还没认识的量子引力定律能告诉我们，这些奇点的行为如何，会对附近的时空做些什么，它们的作用是否会对我们生活的宇宙产生巨大的影响？因为裸奇点（如果能够存在的话）可能强烈影响我们的宇宙，我们非常想知道宇宙监督猜想是不是

正确的，量子引力定律对奇点行为会有什么预言。想弄清这些问题，不会很快，也不会太容易。

[1]　$10^{-43}$秒是普朗克-惠勒时间。它由公式$\sqrt{G\hbar/c^5}$近似给出。注意，这个时间是普朗克-惠勒面积（12章）的平方根除以光速。（公式中的符号和数值在普朗克-惠勒面积的脚注中已经说明了。）[35]

[2]　以上的描述是以惠勒-德维特、霍金-哈特尔建立量子引力定律的方法为基础的，尽管他们的方法不过是现在人们正在研究的众多方法的一种，但我认为它成功的希望更大一些。

[3]　赌约写的是："鉴于S·W·霍金诚信裸奇点为可诅咒者，应为经典物理学所禁戒；而J·普雷斯基尔与K·索恩以裸奇点为量子引力客体，能不为视界所隐藏而令全宇宙都能看见。故两方约定，霍金以100英镑对普雷斯基尔-索恩50英镑赌：在平直时空不可能产生奇异的任何形式的经典物质或场，在通过经典爱因斯坦方程与广义相对论相联系时，结果也不可能是裸奇点。输家向赢家提供蔽体的衣服，衣服上须绣适当的认输字据。（签名，霍金按的手印，时间是1991年9月24日。）——译者"

# 第14章　虫洞和时间机器

为了洞察物理学定律，作者问：

高度发达的文明

能在超空间凿开虫洞

做快速星际旅行

并从时间机器回到过去吗？[1]

## 2

上完1984～1985学年的最后一堂课，我坐进办公室的椅子，想好好放松一下。这时，电话铃响了，是我多年的老朋友，康奈尔大学天体物理学家卡尔·萨根（Carl　Sagan）打来的。“基普，打扰了！”他说，“我刚写完一本小说，讲人类第一次同外星文明打交道。不过有点儿麻烦。我想尽量把科学的东西写得准确一些。我怕把某些引力物理的东西弄错了，你能替我看看吗？”我当然愿意。卡尔是个聪明的伙计，那书一定很有意思，而且还可能很逗人。再说，老朋友的请求，我怎么能不答应呢？

几个星期后，小说寄来了。隔行打印的稿子，三英寸半厚的一摞。

我和前妻琳达（Linda）和我们的儿子布里特（Bret）正要去圣克鲁斯看大学毕业的女儿卡丽丝（Kares）。我把书稿塞进旅行包，放

在琳达的野马车后座上，从帕萨迪纳出发了。

琳达和布里特轮流开车，我一边看书一边思考。（他们跟我在一起生活了多年，已经习惯我的这种行为了。）小说很逗人，但卡尔确实有点儿问题。他让他的女主角阿洛维（Eleanor Arro-way）落进地球附近的一个黑洞，然后像图13.4那样穿过超空间，一小时后出现在26光年远的织女星旁。卡尔不是相对论专家，不熟悉微扰计算的结果[3]：不可能从一个黑洞的中心穿过超空间到我们宇宙的另一部分。任何黑洞都不断受电磁真空小涨落和少量辐射的攻击。这些涨落和辐射落进黑洞时，被黑洞引力加速到巨大能量，然后暴雨般落向可能被人们借以穿越超空间的任何“封闭小宇宙”或“隧道”或宇宙飞船。计算不容置疑，任何做超空间旅行的飞船都会在启动前就被“暴雨”摧毁。卡尔的小说得改。

从圣克鲁斯回来，在5号州际公路上弗雷斯诺西边的某个地方，我突然闪出一个念头，也许，卡尔可以把他的黑洞换成穿过超空间的虫洞。

虫洞是宇宙中相距遥远的两点间的一条假想捷径。它有两个洞口，例如，一个在地球附近，另一个在26光年外织女星轨道附近。两个洞口通过超空间的隧道相联结（虫洞），可能只有1千米长。假如我们从地球附近的洞口走进隧道，只经过1千米，就到达另一洞口，出现在（从外面的宇宙看来）26光年远的织女星旁。

图14.1用嵌入图画了这样一个虫洞。与通常的嵌入图一样，在这个图中，我们的宇宙也理想化为二维的，而不是三维的（见图3.2和3.3）。宇宙的空间在图中表现为一张二维面。在纸上爬行的蚂蚁感觉不到纸是平整的还是褶皱的，同样，宇宙中的我们也不太清楚我们的宇宙在超空间里是平直的还是像图那样弯曲的。然而，有一点褶皱也是重要的，这样地球和织女星才可能在超空间里相邻，从而才可能通过很短的虫洞联结起来。空间有了虫洞，我们就和在嵌入图的曲面

上爬行的蚂蚁和小虫那样，有两条可能的从地球到织女星的道路：沿着外面宇宙的26光年的长路和穿过虫洞的1公里的捷径。

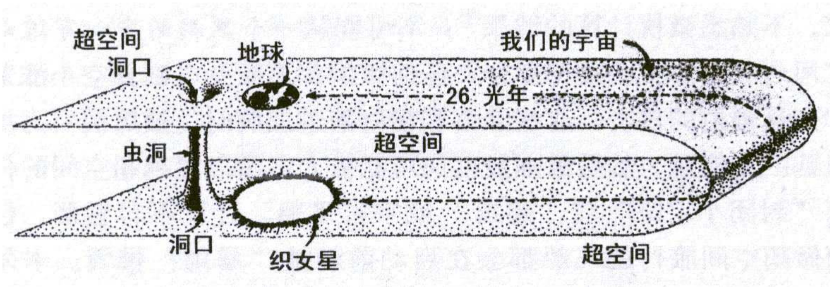



图14.1　通过超空间连结地球和26光年外的织女星的1公里长的虫洞。

假如虫洞在地球上，那么洞口在我们面前像什么样子呢？在嵌入图的二维宇宙中，洞口画成了圆，因此在我们的三维宇宙里，它应该是圆的三维表象，也就是一个球。实际上，洞口可能有点像无旋转黑洞的球状视界，不过有一个重要的区别：黑洞的视界是“单向”曲面，任何事物都能进去，但没有东西可以出来。而虫洞口是“双向”曲面，我们能从两个方向穿过它，可以走进洞里，也可以回到外面的宇宙。向球状洞口内看，可以看见来自织女星的光。光从织女星附近的洞口进入虫洞，像穿过光导管和光纤那样穿过它，然后从地球的洞口穿出来，射进我们的眼睛。

虫洞不仅是科幻小说家凭空想象的东西，早在1916年就从数学上在爱因斯坦场方程的解里发现它了。[1]那时，爱因斯坦的场方程刚建立几个月。后来，在50年代，惠勒和他的研究小组又用不同的数学方法对它们进行过广泛的研究。不过，在我1985年在5号公路旅行以前，所发现的那些作为爱因斯坦方程的解的虫洞，没有一个适合于萨根的小说，因为没有谁能够安全穿越它们。它们每一个都随时间奇怪地演变：虫洞在某个时刻产生，短暂地打开，然后关闭、消失一从产生到消失，时间极短，没有事物（人、辐射或任何形式的信号）能在这么短的时间内从一个洞口穿过它到达另一个洞口。谁想去试试，一

定会在它的消失中毁灭。图14.2画了一个简单的例子。

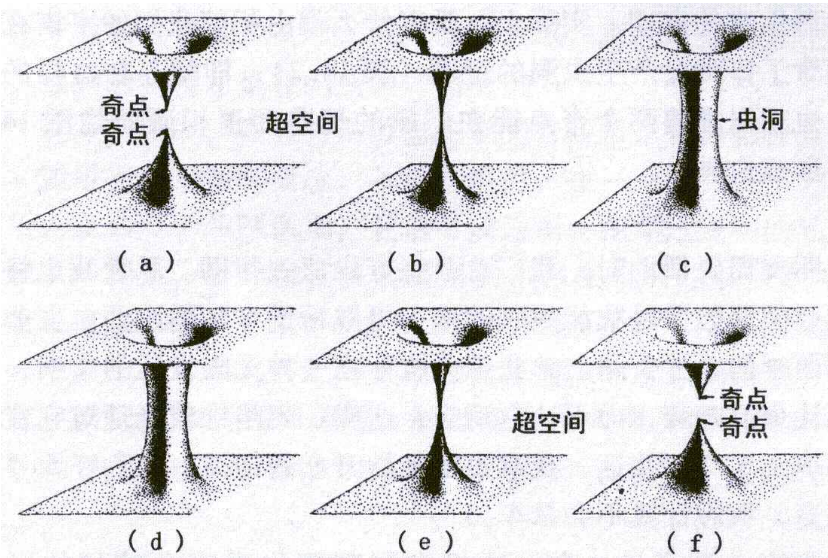



图14.2　洞内无任何物质的完全球状虫洞的演化。（这个演化过程是普林斯顿大学惠勒的年轻助教克鲁斯卡（Martin Kruskal）在50年代中期从爱因斯坦场方程的解中发现的。）初始时（a），没有虫洞，在地球和织女星附近各有一个奇点。然后，在某一时刻（b），两个奇点在超空间里生长、相遇，然后湮灭，在湮灭中生成虫洞。虫洞周长在（c）增大，然后又收缩（d），最后消失（e），产生两个奇点（f），就像虫洞产生前的样子——但有一点决定性的不同：初始奇点（a）像大爆炸，时间从它流出，它也能生成某些事物：大爆炸产生宇宙，初始奇点产生虫洞。而最后的奇点（f）不一样，它像大收缩（第13章），时间流进它，万物被它毁灭：大收缩毁灭宇宙，它毁灭虫洞。任何企图在虫洞打开的短暂时间里穿过去的事物，都将在虫洞关闭时被捕获，随它自身一起消失在最后的奇点（f）。[2]

几十年来，我和大多数物理学同行一样，也在怀疑虫洞。照爱因斯坦场方程的预言，虫洞的寿命本来就很短暂，在辐射的随机打击下还会更短。辐射（根据伊尔德莱（Doug　Eardley）和雷德蒙特（Ian Redmount）的计算）被虫洞引力加速到超高能，虫洞的喉管在强大辐射的轰击下，比以往更快地收缩、关闭——霎那间就完了，仿佛根本就不曾存在过。

还有另一个怀疑的理由。我们知道黑洞是星体演化不可避免的结果（天文学家在我们星系中大量看到的那些大质量的缓慢旋转的恒星

在死亡时会坍缩形成黑洞），但在自然界却没有类似的虫洞生成的方式。实际上，没有什么理由相信我们的宇宙在今天包含了任何会产生虫洞的奇点（图14.2）；即使存在这样的奇点，也难以理解两个奇点能在广阔的超空间里相遇而像图14.2那样形成虫洞。

朋友需要帮助时，我们总会想方设法去帮助。尽管我也怀疑虫洞，但那似乎是我能找到的惟一可以帮助卡尔的东西。在弗雷斯诺西畔的5号公路上，我想大概存在一种无限发达的文明，可以总让虫洞开着，而不让它消失；这样，阿洛维就能通过它在地球和织女星之间往返。我拿出纸笔就开始算起来。（幸好5号公路很直，我做计算不会晕车。）

为使计算容易一些，我把虫洞理想化为完全球状的（图14.1也是这样的，不过三维宇宙在图中压缩成二维，虫洞的截面是圆）。接着，我以爱因斯坦场方程为基础，做了两页计算，发现三件事情：

第一，保持虫洞开放的惟一方法是，用某种类型的物质贯穿虫洞，靠引力作用将洞壁撑开。我把这种物质称为奇异的，因为下面会看到，它与人类所见过的任何物质都大不一样。

第二，我发现，奇异物不仅像要求的那样会把洞壁向外推，而且当光束通过时，它还会凭引力将光线外推，使光束分离。换句话说，奇异物像一个“散焦镜”，靠引力将光束分开。见卡片14.1。

第三，我从爱因斯坦场方程知道，为了靠引力让光束分散，靠引力将虫洞壁撑开，贯穿虫洞的奇异物在光束看来必须具有负能量密度。这需要解释一下。想一想，引力（时空曲率）由质量产生（卡片2.6），而质量与能量等价（卡片5.2，等价性体现在爱因斯坦的著名方程$E=Mc^2$）。就是说，可以认为引力是能量产生的。现在，我们从光束的角度——也就是从某个以（近）光速穿越虫洞的观测者的角度——来计算虫洞内物质的能量密度（每立方厘米的能量），然后沿光

束轨迹求它的平均。结果，只有在平均能量密度为负时，光束才能分散，虫洞才会张开——这样，虫洞的物质才是我们所谓“奇异的”。[4]

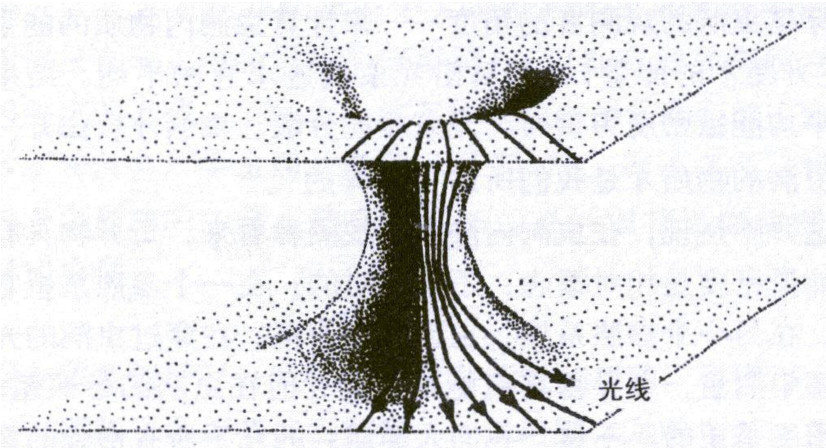
这并不是说，在虫洞内静止的观测者看来，奇异物具有负能量。能量密度是相对概念，不是绝对的；在一个参照系里它可以为负，在另一个参照系里，它也可以为正。在穿过虫洞的光束的参照系中测量，奇异物有负能量密度；但在虫洞的参照系测量，能量密度是正的。不过，我们人类遇到的几乎所有形式的物质在每一个参照系中都具有正的平均能量，物理学家长期以来一直怀疑奇异物的存在。我们猜想，物理学定律大概严禁这样的奇异物，但一点儿也不清楚它们是如何做到这一点的。

卡片14.1　让虫洞打开：奇异物

任何球状虫洞都将分散穿过它的光束。为看清这一点，想象（如图所示）光束在进入虫洞前经过一会聚透镜，这样光线沿径向向虫洞中心会聚，然后，光线继续沿径向穿行（它们如何还能运动呢？），就是说，在从另一洞口出现时，它们沿径向散开，像图中那样离开虫洞中心。光束就这样解散了。

令光束解散的虫洞的时空曲率，是贯穿虫洞并使它张开的“奇异”物产生的。而时空曲率等价于引力，所以实际上是奇异物的引力让光束散开的。换句话讲，奇异物排斥光束的光线，把它们从它自己身边赶走，从而它们也相互分离散开了。

这与引力透镜发生的事情正好相反（图8.2）。在那儿来自遥远恒星的光被途中的恒星或星系或黑洞的引力所吸引、聚焦；在这里，光却被散焦了。




我在5号公路上想，也许我们对奇异物存在的偏见是错误的。也许奇异物能够存在。这是我能发现的惟一可以帮助卡尔的。所以，回到帕萨迪纳，我就给卡尔写了一封长信，向他解释，为什么他的女主角不能借黑洞做星际旅行。我建议让她去穿过虫洞；小说中还应该有某个人发现奇异物真能存在，而且可以用来打开虫洞。卡尔愉快地采纳了我的建议，写进了最后的定稿。那小说叫《接触》（Contact）。5

给卡尔的信寄出后，我突然想，他的小说可以作为学生学广义相对论的教学工具。1985年秋，莫里斯（Mike　Morris，我的学生）和我为了帮助这些学生，开始写一篇论文，关于奇异物支撑的虫洞的广义相对论方程和这些方程与萨根小说的联系。

我们写得很慢，其他更急迫的事情赶到前头去了。1987～1988年的冬天，我们把稿子交给《美国物理学杂志》。[3]还没发表，临近博士毕业的莫里斯正申请博士后研究，他在申请书里附上了我们的文章。帕奇（宾夕法尼亚州立大学教授，我和霍金以前的学生）收到了申请，读了我们的稿子后给莫里斯写了封信：

“亲爱的麦克，……据霍金和埃利斯书中的命题9.2.8，加上爱因

斯坦场方程，立刻就能得到，任何虫洞[都需要奇异物来支撑]……您忠实的D.N.帕奇。”

我觉得自己太傻了。我从没深入学过整体方法6（霍金和埃里斯一书的主题），[4]现在付出代价了。我在5号公路上不太费力地得出，为了打开完全球状的虫洞，需要奇异物的贯通。现在，帕奇用整体方法更不费力就得到，打开任何（球状的、立方体状的或有任意形变的）虫洞，都必须有奇异物穿过。后来我听说，甘农（Dennis Gannon）和C·W·李在1975年得到过几乎相同的结论。

虫洞需要奇异物打开的发现，在1988～1992年间激起了理论研究热潮，中心问题是，“物理学定律容许奇异物存在吗？如果是的，那应在什么条件下呢？”

解开这个问题的钥匙，霍金在70年代就已经准备好了。1970年，霍金在证明黑洞面积总会增加时（第12章），不得不假定任何黑洞视界附近不存在奇异物。假如视界边有奇异物，霍金的证明就失败了，他的定理将失去意义，视界面积可以收缩。

然而，霍金并不太替这种可能担心。看来，在1970年大家都愿意相信奇异物不可能存在。

可是，1974年出现了令人大吃一惊的事情：霍金从他黑洞蒸发（第12章）的发现中顺便推测，黑洞视界附近的真空涨落是“奇异的”：[5]从视界附近流出的光束看，它们具有负平均能量密度。事实上，令黑洞在蒸发中收缩从而违背霍金面积增加定理的，正是真空涨落的这种奇异特性。由于奇异物对物理学太重要了，我还是好好解释一下：

回想一下卡片12.4讨论的真空涨落的起源和本质：当我们试图将电场和磁场从某个空间区域拿走，也就是当我们想产生理想真空时，

总会留下一些随机的不可预测的电磁振荡——由相邻空间区域的场之间的“交流”产生的振荡。“这里”的场向“那里”的场借走能量，给“那里”的场留下能量亏损，即在那里出现瞬间负能量。然后，那里的场立刻收回能量，还附带着一点盈余，使自己拥有瞬间正能量。这样的过程，一直不断地进行着。

在地球的正常情况下，这些真空涨落的平均能量为零。能量处在盈亏状态的总时间相等，所以平均说来没有盈亏。而霍金1974年的计算意味着，在蒸发黑洞的视界附近会出现不同的情况。视界旁的平均能量，至少在光束看来一定是负的，就是说，真空涨落是奇异的。

这些事情是怎样发生的？具体情况到80年代初才有结果。那时，宾夕法尼亚州立大学的帕奇、牛津的康迪拉斯（Philip Candelas）和其他许多物理学家用弯曲时空的量子场定律广泛深入地研究了黑洞视界对真空涨落的影响。他们发现，视界的影响是关键。视界使真空涨落扭曲，出现地球上没有的形状。通过扭曲，平均能量密度成为负的，这样，真空涨落也成为奇异的了。

真空涨落在什么条件下变奇异呢？它们能在虫洞内表现奇异特性而令虫洞打开吗？帕奇发现奇异物质是打开任何虫洞的惟一途径，这两个问题是对他的发现所激起的研究潮流的巨大冲击。

答案来之不易，而且也不彻底。克林卡莫（Gunnar Klinkhammer，我的学生）证明，在平直时空，即在远离一切引力物体的地方，真空涨落不可能是奇异的——它们不可能具有光束看到的负平均能量密度。另一方面，瓦尔德（惠勒以前的学生）和尤泽维尔（Ulvi Yurtsever，我以前的学生）证明，在弯曲时空的很多情况下，曲率会扭曲真空涨落从而使它们成为奇异的。[6]

虫洞想脱离这样的环境吗？虫洞的曲率能通过扭曲作用让真空涨落成为奇异的从而打开虫洞吗？在这本书出版时，我们还不知道。

1988年初，奇异物的理论研究方兴未艾时，我才发觉萨根的电话所激起的那些研究是多么有力。在实验家可能会做的所有真实物理实验中，最可能为物理学定律带来深刻新认识的是那些最猛烈推进定律的实验；同样，当理论家在探索超越了现代技术的物理学定律时，在他可能考察的所有思想实验中，最可能产生深刻新见解的是动力最强的。但所有这些思想实验对物理学定律的推动，都不如萨根给我的电话触发的那一个——它问，“物理学定律容许无限发达的文明做些什么？又严禁他们做什么？”（所谓“无限发达的文明”说的是他们的能力只受物理学定律的限制，而不存在行为方式、工作技巧等任何其他事物的局限。）

我相信，我们的物理学家总想回避这样的问题，因为它们太像科幻小说了。虽然我们很多人都喜欢读科幻小说，甚至还写一些，但我们怕同行笑话在科幻小说的边缘做研究。于是，我们更愿意研究另外两个不那么“幻想”的问题：“宇宙中哪些事情会自然发生？”（例如，黑洞自然出现吗？虫洞自然出现吗？）“我们人类凭现在和不远将来的技术能做些什么？”（例如，我们能生产像钚那样的新元素来造原子弹吗？我们能制造高温超导体来降低悬浮列车和超大粒子对撞机的超导磁体的费用吗？）

我在1988年才明白，我们物理学家在这些问题上原来是相当保守的。那时，已经有一个萨根式问题（我愿意这么叫）开始有结果了。我们问，“无限发达的文明能为快速星际旅行留住虫洞吗？”莫里斯和我认定奇异物是留住虫洞的关键，而且，为了认识在什么条件下物理学定律允许（或不允许）奇异物存在，我们也激发了多少有些结果的研究。

假如我们的宇宙在大爆炸中诞生时完全没有虫洞，那么，亿万年以后，当智慧生命创造出（假想的）无限发达的文明时，那个无限发达的文明能为快速的星际旅行构造虫洞吗？物理学定律允许在原来没有虫洞的地方构造虫洞吗？允许我们的宇宙空间发生这样的拓扑改变

吗？

这些问题是萨根星际旅行问题的后一半；前一半问题是，如何留下造好的虫洞。萨根通过奇异物把它留下了。后一半问题在他的小说里却悄悄溜过了。他描绘说，阿洛维旅行的虫洞现在是靠奇异物留下的，但它是在遥远的过去由某个无限发达的文明创造的，关于他们的所有历史记录都失去了。

我们物理学家当然不愿意把虫洞的产生推给史前文明，我们想知道，宇宙的拓扑在物理学定律限制下，现在能否改变？怎么改变？

我们可以设想两个在原来没有虫洞的地方构造虫洞的方法：一个是量子方法，一个是经典方法。

量子方法依赖于引力真空涨落（卡片12.4），也就是类似于上面讨论的电磁真空涨落的引力现象：相邻空间区域的能量“借贷”往来引起的空间曲率的随机的概率涨落。一般认为，引力空间涨落是处处都有的，但在普通条件下它们太小了，还没有被实验探测到。

当电子被限制在越来越小的区域时，它们的随机简并运动会越来越强（第4章），同样，引力真空涨落在小区域比在大区域强，也就是短波长的涨落比大波长的强。1955年，惠勒以原始粗略的方式结合量子力学和广义相对论的定律，得出在普朗克-惠勒长度，$1.62\times10^{-33}$厘米或更小的区域内，存在着巨大的真空涨落，如我们所知，那空间“沸了”，成了一堆量子泡沫[7]——也就是构成时空奇点的那种量子泡沫（第13章；图14.3）。[7]

于是，量子泡沫无处不在：在黑洞内部，在星际空间，在你屋里，在你头脑中。但是，要看量子泡沫，必须拿（假想的）超级显微镜去看越来越小的空间和空间里的东西。从你我的尺度（100多厘米）看到原子（$10^{-8}$厘米）、原子核（$10^{-13}$厘米），这样看下去，再

小$10^{20}$，直到$10^{-33}$厘米。先看到的“大”尺度空间是完全光滑的，只有一定的（小小的）曲率。然而，在接近、经过$10^{-32}$厘米时，我们会看到空间开始卷曲缠绕了，先很缓和，然后越来越强烈，当$10^{-33}$厘米大小的区域完全走进超级显微镜的目镜时，空间已经成了一团概率的量子泡沫。

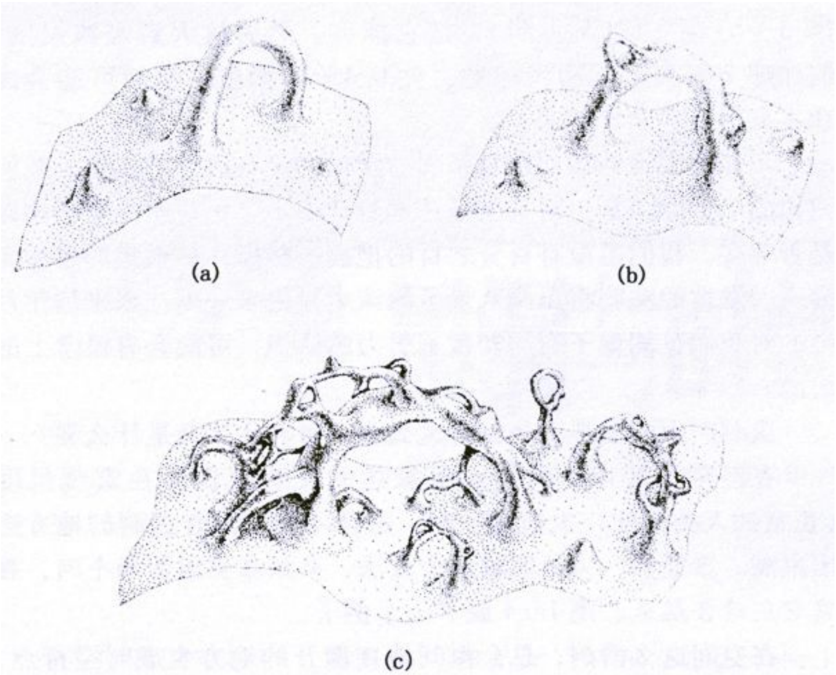


图14.3　（同图13.7）量子泡沫的嵌入图。空间的几何与拓扑是不确定的，而是概率性的。例如，对于如图所示的（a）的形态，它有0.1%的概率，（b）为0.4%，（c）为0.02%，等等。

因为量子泡沫处处都有，我们不禁会想象让某个无限发达的文明走近量子泡沫，找出一个虫洞（例如，有0.4%概率的图14.3（b）中的“大”洞），把它抓住，然后放大到经典尺度。假如那文明真是无限发达的，凭0.4%的概率，他们可能会成功，真的会吗？

不知道，因为我们对量子引力定律还没有很好的认识。我们无知的一个原因是，对量子泡沫本身认识不够，甚至，量子泡沫是否存

在，我们也没有百分之百的把握。然而，萨根式的思想实验——发达的文明将虫洞从量子泡沫中拉出来——在未来的年月里，对我们巩固量子泡沫和量子引力的认识，可能会有概念上的帮助。

虫洞产生的量子方法就讲这么多。经典方法又是什么呢？

在经典方法中，我们无限发达的文明应设法在宏观尺度（正常的人类尺度）上扭曲空间，这样才能在没有虫洞的地方造出虫洞。很显然，为了实现这个方法，必须在空间凿两个洞，再将它们缝合起来。图14.4画了一个例子。

在空间这么凿洞，总会瞬间地在凿开的地方生成时空奇点，也就是时空终结的尖点，而奇点是与量子引力相关的东西，所以这样的虫洞制造方法，实际上还是量子力学的，而不是经典的。在认识量子引力定律前，我们不会知道这种方法是否可行。

没有出路了吗？难道说，造虫洞的方法都得与我们还没认识的量子引力定律纠缠——而没有完全的经典方法吗？

有，但多少有些奇怪——而且得付出很大的代价。1966年，格罗赫（惠勒在普林斯顿的学生）用整体方法证明，通过时空光滑的无奇点扭曲，我们能够构造一个虫洞，但在构造中，不论从什么参照系看，时间也被扭曲了。[8]更具体地说，在构造虫洞的过程中，既可沿时间向前，也能向后；[8]不论造洞的是什么“机械”，它的作用都必然像一台时间机器，带着东西从后来的时刻回到以前的时刻（但不能回到开始造虫洞以前）。

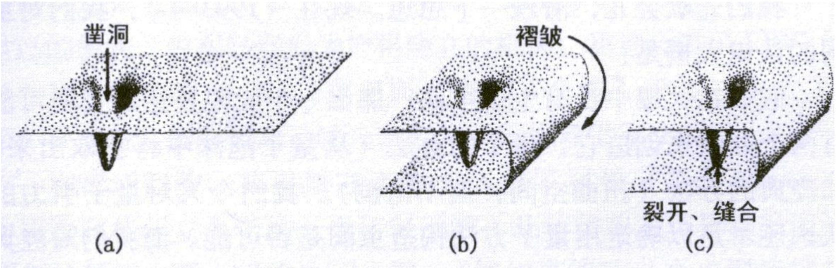



图14.4　造虫洞的一种方法。（a）在空间曲率上凿出一个洞。（b）洞外的空间在超空间中缓慢褶皱。（c）在那个洞的尖端凿一个洞，在洞下面的空间也凿一个洞，然后将两个洞的边缘“缝合”起来，初看时，这个方法是经典的（宏观的），然而，凿开的洞至少会瞬时产生与量子引力定律相关的时空奇点，所以这个方法实际上也是量子的。

1967年，对格罗赫定理的普遍反应是，“物理学定律肯定会禁止时间机器，所以，用经典的方法，也就是不在空间凿洞，是不可能造出虫洞来的。”

在以后的十几年里，我们过去认为肯定的事情看来是错了。（例如，我们在1967年怎么也不会相信黑洞会蒸发。）这告诫我们应当谨慎。为了谨慎，也因为萨根式问题的激发，我们在80年代后期开始提出这样的问题：“物理学定律真的严禁时间机器吗？如果是的，它如何去禁止呢？这些定律会以什么方式维护这样的禁令呢？”下面我还将回到这个问题。

我们先歇会儿，清理一下思想。现在（1993年），我们对虫洞的认识大概是：

假如在大爆中没有生成虫洞，那么一个无限发达的文明可能有两个办法来创造它，量子的办法（从量子泡沫中将它取出来）和经典的办法（扭曲空间，但不凿洞）。我们今天对量子引力的认识还不足以确定用量子方法构造虫洞是否可能。而我们对经典引力定律（广义相对论）的足够认识则确实令我们相信，用经典方法构造虫洞是允许的，但是不论构造者是什么“机械”，时间在所有参照系看来都会被它强烈扭曲，结果，它（至少在短时间内）成了一台时间机器。

我们还知道，假如无限发达的文明凭某个方法得到了一个虫洞，那么，令虫洞打开（这样可以用来做星际旅行）的惟一办法是，让奇异物穿过洞。我们知道电磁场的真空涨落很有可能是一种奇异物：在很多不同的情况下，它们在弯曲时空里都可以表现出奇异性（在光束看来，具有负平均能量密度）。然而，我们不知道它在虫洞内是否还能奇异，从而为我们把洞打开。

在接下来的几页里，我假定某个无限发达的文明已经通过某种方法获得了一个虫洞，而且靠某种奇异物让洞一直开着；我的问题是，除了星际旅行外，这个文明还可能用虫洞来做些什么。

[1] 这一章主要是照我个人的观点写的，所以不像其他章节那么客观；而且对别人的研究讲得很少，很不全面。

[2] 这一章主要是照我个人的观点写的，所以不像其他章节那么客观；而且对别人的研究讲得很少，很不全面。

[3] 见第13章“最佳猜想”一节。

[4] 用专业术语说，奇异物“违背了弱平均能量条件”。

[5] 特别请看萨根《接触》第347、348和406页。那儿的奇异物条件（在穿过虫洞的光束看来，奇异物有负平均能量密度）的表述不同，但是等价的：从某个静止在虫洞里的人看来，奇异物一定在径向上有比能量密度还大的张力。

[6] 第13章。（霍金和埃利斯的那本书即《时空的大尺度结构》，是用整体微分几何方法写的一部广义相对论专著。很遗憾，我不能将命题9.2.8用几行通俗文字说明白。——译者）

[7] 普朗克-惠勒长度是普朗克-惠勒面积（原来出现在黑洞熵公式中，见第12章）的平方根，公式为 $\sqrt{G\hbar/c^3}$ ，各符号意义前面注释过了。

[8] 我真想画一个简单明白的图来说明这种光滑的虫洞是如何实现的，遗憾的是我画不出来。

# 1

1986年，第14届半年度的德克萨斯相对论天体物理学会议在伊利诺斯的芝加哥举行，从1963年在德克萨斯达拉斯第一次讨论类星体（第7.9章），这一系列“德克萨斯会议”就具有了自己的模式，现在已经成为严格建立的机构。我到会讲了LIGO的梦想和计划（第10章），莫里斯（我的“虫洞”学生）也去了，第一次出现在国际相对论物理学家和天体物理学家面前。

在讲话间隙，莫里斯在走廊上认识了罗曼（Tom Roman），中康涅狄格州立大学的一个年轻助教，几年前曾对奇异物发表过深刻的见解。两人很快谈到虫洞。“假如真能让一个虫洞持续打开，那么它会允许在星际距离间的旅行比光速还快。”罗曼指出，“这是不是说，我们也能借虫洞反时间旅行呢？”

麦克和我觉得自己真笨！当然，罗曼是对的。事实上，我们在儿童时代就从一首有名的滑稽诗里听到过这样的时间旅行：[2]

女孩儿呀，贝蕾

来去呀，光难追。

相对论呀，捷径，

今日出门呀，

昨夜回。

在罗曼和这首小诗的激发下，我们明白了如何用两个彼此相对以光速运动的虫洞来建一台时间机器。[3]（这种时间机器有点儿复杂，我不准备在这儿讲；我很快会讲另一种更简单、更容易描述的时间机

器。）

我喜欢孤独，喜欢一个人去山里，去远离尘嚣的海边，甚至躲进小屋去思考。新思想总是从长时间安静的没有惊扰的孕育中慢慢产生出来的；大多数必须进行的计算也是经过好多天或者好多个星期的持续紧张的全神贯注的活动才能实现的。一个突然的电话也能令我分心，耽误几个小时。于是，我藏起来了。

但躲得太久也不是好事。我时刻需要与不同观点和专长的人交流，从与他们的对话中得到灵感。

到现在，我在本章已经讲了三个这样的例子。如果卡尔不打电话来让我从科学的角度为他改小说，我永远不会去研究虫洞和时间机器；如果没有帕奇那封信，莫里斯和我不会知道无论什么形状的虫洞，都需要奇异物来打开；还有，如果没有罗曼的证明，莫里斯和我大概还不知道，发达的文明可以很容易地通过虫洞制造时间机器。

接下来我再讲几件给我带来巨大灵感的事情。当然，并不是所有思想都是这样产生的，有的还是通过自己的沉思得到的。

1987年6月初，几个月的课讲完了，几个月和我的小组以及LIGO计划在一起的日子也结束了，我疲惫不堪，一个人躲了起来。

那年的整个春天，总有件事情在困扰着我，我想先不去理它，等安静下来再去考虑。现在，宁静的日子终于来了。一个人时，困惑从潜意识浮现出来，我开始检验它：“时间在通过虫洞时如何决定它自己的连结方式？”这是问题的要害。

为把问题说得更具体些，我想了一个例子：假定我有一个很短的虫洞，它的隧道在超空间里只有30厘米，两个洞口（即两个球）的直径为2米——把它放在帕萨迪纳我的家里。我从洞爬过去，自己觉得很快就从另一端出来了，没有一点耽误；事实上，我的头爬出第二个

洞口时，脚还留在第一个洞口的外面。这似乎意味着，坐在屋里沙发上的妻子卡洛丽会看到，我的头从第二个洞口露出来时，我的脚正在往洞里爬，即图14.5的样子，真会这样吗？如果是的，那么时间在“穿越虫洞”和在虫洞外面的“连接方式”是一样的。

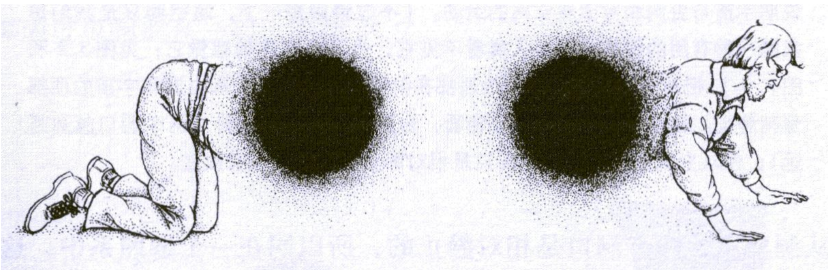


图14.5　我在超空间中爬过一个短虫洞。

另一方面，我也问自己，虽然我自己觉得几乎没花什么时间就穿过了虫洞，但卡洛丽也许会等一个小时才看见我从第二个洞口爬出来，可能这样吗？当然，也许她在我爬进去的一个小时前就看见我出来了，这是不是也可能呢？假如是这样，那么时间在穿越虫洞和在虫洞外面的连结方式就不一样了。

什么事情能让时间表现得如此怪异？我问自己。反过来，我想，它为什么不应该这样呢？只有物理学定律知道答案。不论怎样，我都应该从物理学定律发现时间到底是如何表现的。

为帮助大家理解物理学定律如何决定时间的连结方式，我构想了一个更复杂的情形。让虫洞的一个出口静止在我的房间里，另一个在星际空间，以光速离开地球运动。虽然两个洞口在相对运动，我们还是假定洞长（通过超空间的隧道长度）总是固定在30厘米。（图14.6解释了为什么当从外面的宇宙看到两个洞口在相对运动时，虫洞还可能保持固定的长度。）于是，从外面的宇宙看，两个洞口处在不同的参照系中，那两个参照系在高速地相对运动着；因此，洞口一定经历着不同的时间流。另一方面，从洞里看，两个洞口是相对静止的，所

以同在一个参照系中，这意味着洞口一定经历着相同的时间流。从外面看，两个洞口经历着不同的时间流；从里面看，却是同一个时间流，怎不令人糊涂！

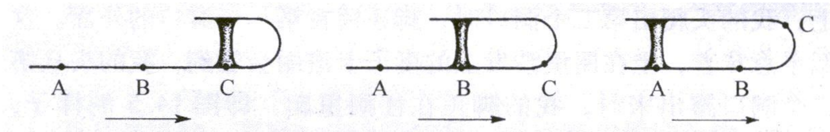



图14.6　为什么在外面的宇宙看到两个洞口在相对运动时，虫洞还能保持固定的长度。每幅图都是图14.1那样的嵌入图，这里画的是剖面。这是一幅快照，说明宇宙与虫洞相对于超空间的运动。（不过请回想一下，超空间只是我们想象的一种有用的假想空间，人类看不见它，也不能实在地感觉它；见图3.2和图3.3。）相对于超空间，宇宙的底部在向图的右方滑行，而虫洞和宇宙的顶部保持静止。相应地，从我们的宇宙看，虫洞口在相对运动着（两个洞口越离越远）；但从虫洞里面看，两个洞口是相对静止的，洞长没有改变。

我一个人静静地想，慢慢地明白了，广义相对论明确预言了两个洞口的时间流，也明确预言了这两个时间流从虫洞比较是一样的，而从洞外比较则是不同的。从这个意义说，如果两个洞口在相对运动，那么时间通过虫洞的连结方式与通过外面宇宙的连结方式是不同的。

我后来发现，不同的时间连结方式暗示我们，无限发达的文明可以用一个虫洞来造时间机器，而用不着两个虫洞。怎么做呢？假如我们无限发达，那是很容易的。

为说明这一点，我还是来讲一个思想实验，人类在实验中是无限发达的生命。卡洛丽和我找到一个很短的虫洞，我们把一个洞口放在家里的起居室，另一个洞口放在门前草地上的家庭飞船。

这个思想实验将告诉我们，时间通过任何虫洞的连结方式，实际上依赖于虫洞过去的历史。不过，为简单起见，我假定在卡洛丽和我得到虫洞时，它有最简单的时间连结方式：通过虫洞内部和通过外面宇宙的连结方式一样。换句话说，假如我爬过虫洞，卡洛丽、我和地球上的每个人都会认为，我从飞船上的洞口露出来的时刻与从起居室

爬进去的时刻几乎是相同的。

确认通过虫洞的时间确实如此连结以后，卡洛丽和我设计了一个实验：我留在一个洞口的家里，卡洛丽带着另一个洞口乘飞船以极高速度去宇宙旅行，然后回来。在整个旅行中，我们的手都通过虫洞握在一起，见图14.7。

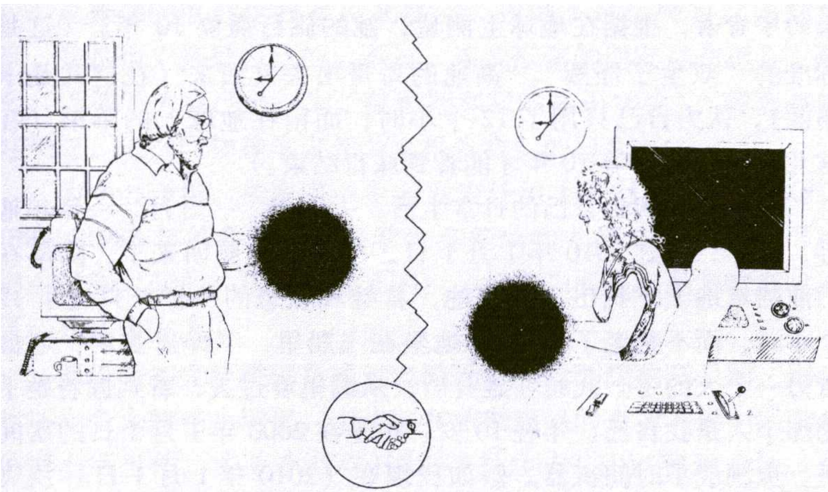


图14.7　卡洛丽和我用一个虫洞构造了一个时间机器。左：我带着一个洞口留在帕萨迪纳的家里，并通过虫洞与卡洛丽握手。右：卡洛丽带着另一个洞口做高速宇宙旅行。中：我们在洞里握在一起的手。

卡洛丽于2000年1月1日上午9:00出发，这个时间是她自己的，也是我的和我们地球上每一个人所测量的。卡洛丽以近光速离开地球，照她测量的时间，她旅行了6个小时，然后掉头回来，以她的时间看，于出发后12小时回到我们家前院儿的草地。[4]我在虫洞里握着她的手，通过虫洞注视着她的整个旅程。显然，我同意，从虫洞看，她真是在出发12小时后，于2000年1月1日晚上9:00回来的。在晚上9:00，我通过虫洞不仅能看见卡洛丽，还看见在她身后的草地和房子。

9点零1分时，我抬头望窗外——只看到空空的草地，没有飞船，

没有卡洛丽和另一个洞口。假如有一台很好的指向窗外的望远镜，我会看见卡洛丽的飞船还在远离地球的航行中。从洞外面的宇宙看，根据在地球上测量，她的旅行需要10年。（这是标准的“双生子怪圈”。[5]高速的哥哥出去又回来（在这儿是卡洛丽），认为自己只用了12个小时；而留在地球上的弟弟（在这儿是我）却得等10年才能看到旅行结束。）

于是，我回到自己的日常生活，一天天、一月月、一年年地等，终于，等到2010年1月1日，卡洛丽远航回来了，降落在门前的草地上。我出去迎接她，看她和预想的一样，只过了12个小时，而不是老了10年。她坐在飞船里，手伸进虫洞，还握着另一个人的手。我站在她身后，从洞里看过去，看到握着她手的那个人是我自己，年轻10岁，正坐在2000年1月1日的房间里。虫洞成了时间机器。假如我现在（2010年1月1日）从飞船的这个洞口爬过去，那么我会在2000年1月1日从屋里的那个洞口出来，与年轻的自己相会。同样，假如年轻的我爬进屋里的洞口，他会在2010年1月1日从飞船的洞口出来。从一个方向穿过虫洞我会年轻10岁；从另一个方向穿过虫洞，我会老10年。

但是，不管是谁，都不可能靠虫洞回到2000年1月1日晚上9点以前，不可能退回到虫洞成为时间机器以前。

广义相对论定律是不容置疑的。假如虫洞能被奇异物打开，那么广义相对论就会预言这些结果。

1987年夏，大约在我从广义相对论得到那个结果1个月以后，里查德·普赖斯给卡洛丽打来电话——他是我的亲密朋友，16年前曾证明黑洞会辐射掉所有的“毛”（第7章）；听说我在研究时间机器，他很担心，怕我疯了或老了，或者……卡洛丽要他放心，我还好好的。

里查德的电话令我有点儿震动，我倒不是怀疑自己头脑糊涂，我

是很少怀疑自己的。不过，连我亲密的朋友都在担心，那么（即使不为自己想，为了莫里斯和我的其他学生），我真要好好想想，怎么向物理学家和公众报告我们的研究。

为小心谨慎，我决定不急着发表任何关于时间机器的东西。1987～1988年的冬天，我跟学生莫里斯和尤泽维尔试图尽可能把虫洞和时间的一切事情都弄明白，只有当所有问题都清澈见底了，我才想发表。

莫里斯和尤泽维尔是通过电脑网络和电话跟我联系的，因为我还一个人躲在小屋里，卡洛丽在威斯康星的麦迪逊做为期两年的博士后工作，头7个月（1988年1月～7月）我跟着她，成了她的“男保姆”。我们在麦迪逊租了房子，我把电脑和书桌搬进小阁楼里，多数时间就呆在那儿思考、计算、写作——主要是为了别的项目，也有部分是关于虫洞和时间的。

为了从有经验的反对者那儿得到启发，在与他们的争论中检验我的思想，我每过几个星期都驱车去密尔沃基，与弗里德曼和帕克（Leonard　Parker）领导的一个杰出的相对论研究小组交谈；偶尔也到芝加哥去，访问另一个由钱德拉塞卡、格罗赫和瓦尔德领导的小组。

3月去芝加哥，我又经历了一次震惊。我在那儿搞了次讨论会，讲述我所认识的虫洞和时间机器。会后，格罗赫和瓦尔德问我（主要意思）：“在发达的文明试图将虫洞变成时间机器时，虫洞不会自动毁坏吗？”

为什么？怎么会呢？找不知道。他们向我解释了。用卡洛丽和我的故事来说，他们解释的大意是：卡洛丽正带着飞船上的洞口飞回地球，我带着另一个洞坐在家里。当飞船离地球在10光年以内时，辐射（电磁波）突然能用虫洞做时间旅行：任何一点离开帕萨迪纳以光速

向飞船靠近的随机辐射，10年后到达飞船（从地球上看），进入那儿的洞口，在10年内及时返回（从地球看）；当它从地球上的洞口出现时，原先的它刚开始启程，于是，它与它自己碰头了——不仅在空间里，而且在时空里——强度增加了1倍。另外，每个辐射量子（光子）在旅行中还会因为洞口的相对运动而获得能量的提高（“多普勒效应”式的提高）。

下一次辐射接着从屋里出去，达到飞船，然后从虫洞回来，遇到刚要离开的原先的它，和自己碰在一起，通过多普勒效应增大能量。辐射源源不断地离去，又源源不断地回来，最后变得无限强大（图14.8（a））。

任何一点辐射经过这样的过程后都会生成一束能量无穷的辐射，在两个洞口间的空间中往来。当辐射束通过虫洞时，格罗赫和瓦尔德认为它会产生无限的时空曲率，可能破坏虫洞，从而虫洞成不了时间机器。

我离开芝加哥，恍恍惚惚地驾车开上去麦迪逊的90号州际公路，满脑子都是在两个相对运动着的虫洞口之间飞来飞去的辐射束的图像；我想借图来计算，到底发生了什么事情。我想明白，格罗赫和瓦尔德是对还是错。

快到威斯康星边界时，头脑里的图像清晰出现了。虫洞不会被毁灭。格罗赫和瓦尔德忽略了一个重要事实：辐射束通过虫洞时，虫洞总会像卡片14.1说的那样将它分离。分离的束从地球上的洞口出现时会在空间散开，只有很少一点辐射能走进飞船的洞口然后从虫洞回到地球来与它自己“碰头”（图14.8（b））。

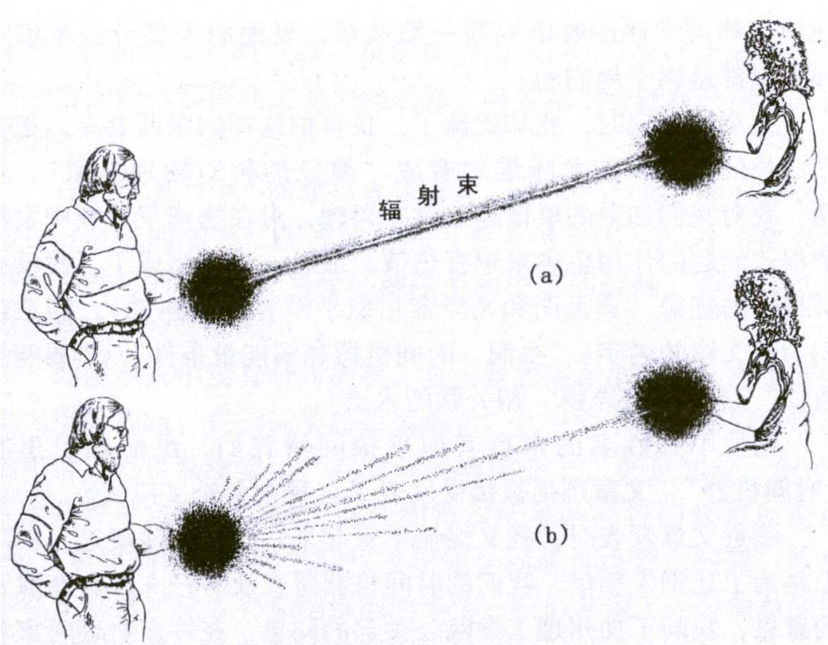



图14.8　（a）格罗赫-瓦尔德提出的虫洞如何可能在成为时间机器前自行毁灭。强烈的辐射束在两个洞口间往来，通过虫洞与自己相遇而加强，最后变得无限强大而毁灭虫洞。（b）实际情况。虫洞使辐射束分散，减少它们相碰的机会；最后的辐射束仍然微弱，不会破坏虫洞。

我一边开车，一边在头脑里“看着”这些辐射叠加。把所有经过虫洞旅行的辐射加在一起（每经过一趟旅行，辐射就分散一些，量越来越小），我发现，最后的辐射束会很弱，远不能破坏虫洞。

结果证明，我的计算是正确的；但后来才知道，我本该更谨慎一些的。虫洞破灭的问题实际上已经在警告我，任何时间机器的制造者都会遭遇意外的危险。

研究生到他们研究的最后一年时，常给我带来巨大的快乐。他们靠自己获得重要发现；在与我讨论时获得胜利；让我学会一些意想不到的事情。莫里斯和尤泽维尔就是这样的两位，我们正在为《物理学评论通讯》写一篇文章，里面的大部分技术细节和思想都是属于他们的。

文章快写完时，我却犹豫了。我害怕这样的东西会令人把正在成长的莫里斯和尤泽维尔看成“疯狂的科幻物理学家”。然而，我对我们知道的事情越来越有兴趣，对在物理学研究中发挥萨根式问题的作用也越来越有热情。最后，论文完成了，我没有讲自己的忧虑（莫里斯和尤泽维尔似乎没有这种感觉），同意他们为论文取的名字：“虫洞、时间机器和弱能量条件”（“弱能量条件”是与“奇异物”相关联的术语）。

两位不知姓名的审稿者似乎很同情我们，虽然题目里有“时间机器”，文章还是被接受发表了。我大大松了口气。

临近文章发表时，我又惴惴不安起来。为了消除疑虑，实际上是为了让别人相信，我们的时间机器研究没有一点儿哗众取宠的意思，我问了加州理工学院公关部的同事。在许多物理学家看来，在大众中故弄玄虚也许是疯狂的行为，而我希望物理学同行们能认真研究的我们的论文。公关部的同事也这样说。

文章发表了，[10]没发生什么事情。正如我所希望的，大众没注意它，但它在物理学家中激发了兴趣，也招来了反对。信一封封飞来，有问问题的，也有挑战结论的。但我们自己的事情已经做完了，有答案了。

朋友们的反应不尽相同。普赖斯还在替我担心，他知道我没疯，也没老，但他怕我坏了自己的名声。苏联朋友诺维科夫是另一种感觉，他着迷了。他正在加利福尼亚圣克鲁斯访问，从那儿来电话说，“我太高兴了，基普！你冲破了阻碍。你能发表时间机器的研究，我也能！”接着，他立刻开始行动了。

[1] 英国小说家Herbert Ceorge Wells（1866-1946）在1895年发表了科幻小说《时间机器》，写一个未来世纪旅行者发现社会分化成了Eloi和Morlocks两个民族。前者曾征服了自然，但不再努力；后者曾被压迫，却成了掠夺者。小说很有名，“时间机器”一词大概是从这儿传下来的。——译者

[2]　这首打油诗是很多年前一个生物学家A.H.R.Buller发表在英国幽默杂志《笨拙》（Punch）上的，不知道有多少相对论的科普读物引用过它。——译者

[3]　这种时间机器和本章后面讲的那些都不能说是人们发现的最早的爱因斯坦场方程的时间机器类解。1937年，斯托库姆（J.van Stockum）发现了一个解，这个解中，一快速旋转的无限长柱体起着时间机器的作用。物理学家从来就认为宇宙间不存在无限长的东西；他们猜测（但没人证明），如果柱体长度有限，它就不会是时间机器。1949年，哥德尔（Kurt Gödel）发现一个爱因斯坦方程的解，描述了一个旋转但既不膨胀也不收缩的全宇宙，一个人只要离开地球到很远的地方然后返回，他就可以到过去旅行。物理学家当然会反驳，他们认为，我们真实的宇宙根本就不像哥德尔的解：它不旋转，至少转得不快；但它却在膨胀。1976年，特普勒（Frank Tipler）用爱因斯坦场方程证明，为了在有限大小的空间区域内造时间机器，必须以奇异物作部分材料。（因为任何可以穿越的虫洞都需要奇异物的贯穿，所以本章描述的以虫洞为基础的时间机器能满足特普勒的要求。）[9]

[4]　实际上，假如卡洛丽要加速到光速并这么快地掉头，她一定会被强大的加速杀死，身体也将被毁坏。不过，这里讲的是物理学家的思想实验的精神，我假定她的身体是高强度材料构成的，能舒适地在加速中生存。

[5]　或者叫“双生子佯谬”（在本书里，我都将paradox译为“怪圈”），在任何一本（狭义）相对论的书里都可以看到对这个现象的描述，但并不能解释；许多书说可以用广义相对论来解释，但似乎也不能令人满意。——译者

# 母子怪圈

在我们的论文激起的抗议中，最有力的是我所谓的母子怪圈[1]：假如我有时间机器（虫洞的或者别的），我就能通过它回到过去，在母亲怀我之前把她杀死，这样就不会让自己出生来害母亲了。[2]

母子怪圈的中心问题是自由意志：作为一个人，我有没有决定自己命运的能力？我真能回到过去杀母亲吗？或者（像多数科幻小说写的那样），当我在她睡梦中举刀的时候，会有什么东西无情地令我住手吗？

即使宇宙中没有时间机器，自由意志现在也是令物理学家手足无措的问题。我们通常总是逃避它，认为它不过是将原本清楚的事情弄得更糊涂罢了。在时间机器问题上，更是如此。所以，在文章发表之前（当然，也在和密尔沃基的同行们认真讨论以后），莫里斯、尤泽维尔和我决定完全回避自由意志问题，坚持不在文章里讨论人类穿越虫洞的事情；我们只谈了一种简单的非生命时间旅行，如电磁波的时间旅行。

文章发表前，我们考虑了很多关于波动通过虫洞回到过去的问题，没有发现在这些波的演化中有什么不可解决的疑惑。最后（也因为弗里德曼的重要启发），我们相信可能不会有解不开的怪圈，[11]在文章里也是这样猜想的。[3]我们甚至还将猜想推广了，认为任何穿过虫洞的非生命物体都不会产生解不开的怪圈。就是这个猜想，引来了强烈的反对。

我们收到的最有意思的一封信，来自奥斯丁德克萨斯大学物理学教授波尔琴斯基（Joe　Polchinski）。他写道，“亲爱的基普，……假如我没理解错，你猜想[在你用虫洞做的时间机器中不会出现解不开的怪圈]。在我看来，似乎……不是这样的。”接着，他巧妙地把怪圈

改成一种简单的形式——从自由意志问题中解脱出来了，于是我觉得可以好好来分析：

拿一个成了时间机器的虫洞，把两个洞口放到行星际空间，相互靠近而且静止不动（图14.9）。现在，从某个恰当的地方以恰当的初始速度向右洞口发射一只台球，球将进入右洞口，沿时间返回，在进入右洞口前（照你我在虫洞外的观察），从左洞口飞出，正好击中原来的自己，从而使它不能进入右洞口回来打自己。

这种情形与母子怪圈一样，都需要回到过去，改变历史。在母子怪圈中，我回到过去，杀了母亲，使她不能生我。在波尔琴斯基怪圈里，台球回到过去，击中自己，使它不能回到过去。

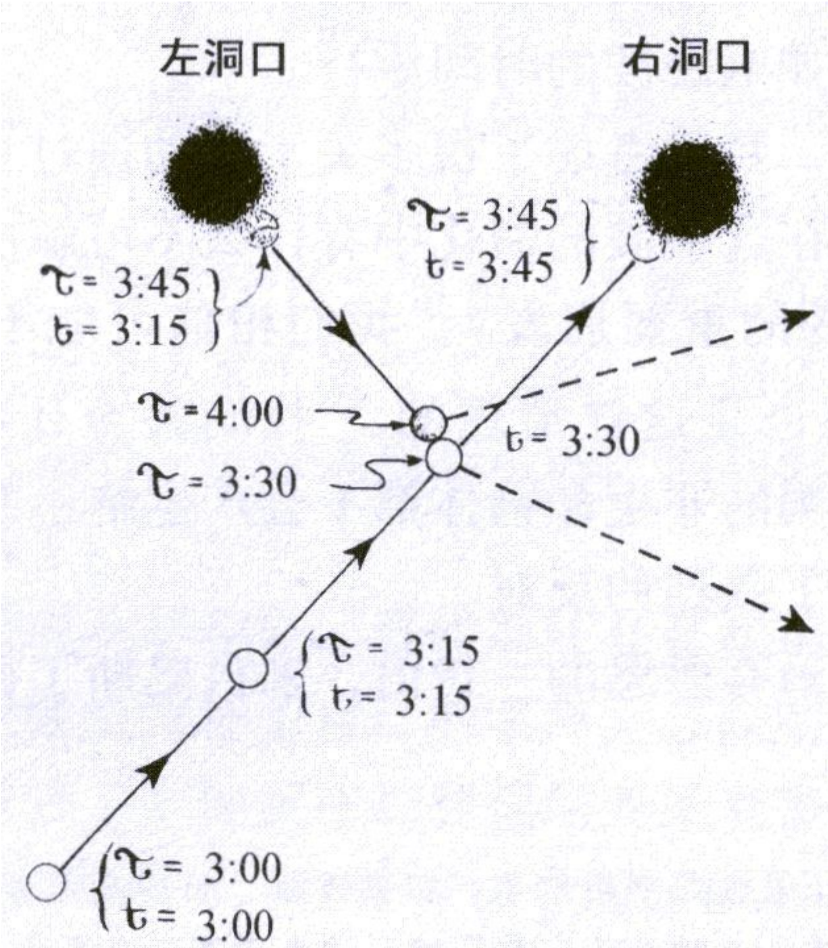



图14.9　波尔琴斯基的台球怪圈。虫洞很短，已成为时间机器。从外面看，进入右洞口的任何事物会在进入30分钟前出现。洞口外的时间流记为t，台球自己经历的时间流记为τ。台球在下午t＝3:00从图示位置发射，速度正好使它在t＝3:45进入右洞口。球从左洞口出现比这早30分，即t＝3:15，然后在t＝3:30击中原先的自己，使它脱离轨道，不能进入右洞，从而不能回来打自己。

两种情形都没有意义。像物理学定律必须逻辑一致一样，由物理学定律所主宰的宇宙演化也应该是逻辑一致的——至少宇宙的经典（非量子力学的）行为应该是这样的；量子力学的行为则更难以捉摸。由于我和台球都是高度经典的事物（也就是说，只有在对我们进行极端精确的测量时，我们才会表现出量子力学行为，见第10章）。

不论我还是台球，都不可能回到过去改变我们的历史。

那么台球到底发生了什么事情呢？为把它弄清楚，莫里斯、尤泽维尔和我集中考察了球的初始条件，即初始位置和速度。我们问自己："在导致波尔琴斯基怪圈的那些初始条件下，是不是还存在别的台球轨迹呢？它们与图14.9不同，但同样是经典台球的物理学定律的逻辑自洽的解。"经过多次讨论，我们认为答案也许是肯定的，但还没有绝对的把握——也没有时间去弄明白了。莫里斯和尤泽维尔博士毕业了，要离开加州，到密尔沃基和特里斯特去做博士后。

幸运的是，加州理工学院的好学生源源不断，又来了两位：埃切维里亚（Fernando Echeverria）和克林卡默（Cunnar Klinkhammer）。他们接过波尔琴斯基的怪圈继续研究：经过几个月断续的数学论证，他们证明，从波尔琴斯基的初始条件出发，确实存在自洽的满足所有经典物理学定律的台球轨道。实际上，存在两条这样的轨道，如图14.10。[13]我将以台球自己的观点依次描述这两条轨道。

在轨道（a）（图14.10左），一只新白球从下午t＝3:00出发，沿着与波尔琴斯基怪圈完全相同的路线（图14.9）向着右边的洞口运动。半小时后，t＝3:30时，这只新的白球被一只看起来旧一些的花球（我们将看到，它是那只球未来的自己）撞在左后边缘。碰撞很轻，新球只稍微偏离了原来的路线，但白球还是被撞成了花球。这只新的花球继续沿着偏离的路线运动，在t＝3:45时进入虫洞口，回到30分钟以前，在t＝3:15时从另一洞口出来。由于路线与波尔琴斯基怪圈的相比发生了偏转，从虫洞出来的变旧了的花球在t＝3:30时从它原来自己的左后边缘轻轻擦过，而不像图14.9那样发生强烈的碰撞和巨大的偏转。这样，球的经历是完全自洽的。

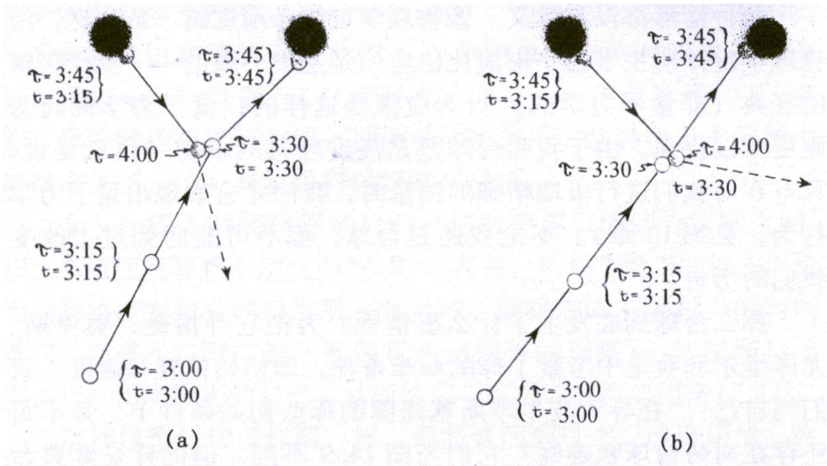



图14.10　波尔琴斯基的母子怪圈（图14.9）的解决：一只在下午3:00以与波尔琴斯基怪圈相同的初始条件（相同的位置和速度）出发的台球可以沿这里的任何一条轨道运动。每条轨道都是自洽的，而且处处满足经典物理学定律。

轨道（b）（图14.10右）与（a）相同，不过球的碰撞方式有些不同，相应地，碰撞的路线也有些不同。特别是，从左洞口出来的旧花球的路线与（a）不同，它沿着这条路线将赶到新球的前头（而不是后面），从它的右前缘（而不是左后边缘）轻轻擦过。

埃切维里亚和克林卡默证明，轨道（a）和（b）都满足台球运动的一切经典物理学定律，因此都可能在真实宇宙中发生（假如真实的宇宙能有虫洞做的时间机器）。

这是最令人不安的。在没有时间机器的宇宙中，这样的情形是永远不会出现的。没有时间机器，一组台球的初始条件只能决定一条而且惟一一条满足所有经典物理学定律的轨道。球只有惟一的运动形式。时间机器把这些都破坏了，现在出现了两种同样合理的球的运动的预言。

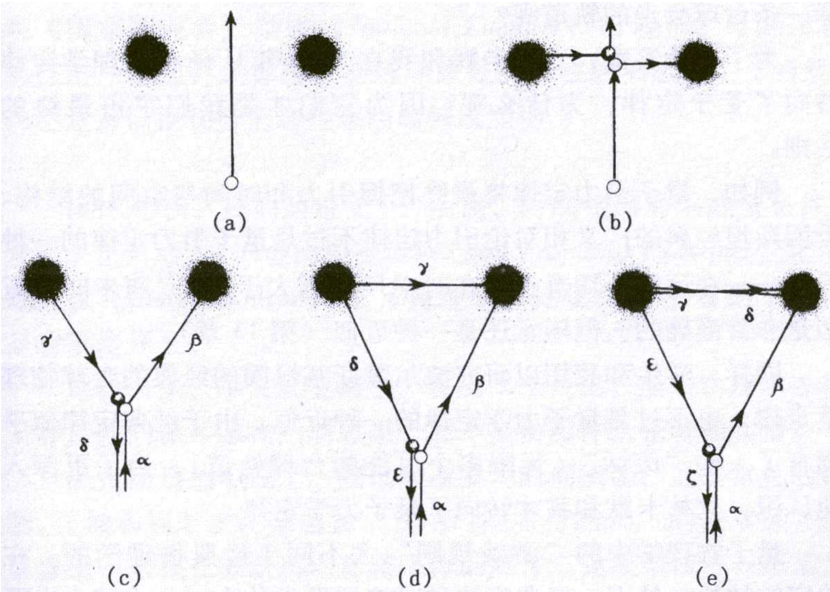
实际上，事情比我们现在看到的更糟：时间机器能为球的运动做出无限多个同样可能的预言，而不只是两个。卡片14.2说明了一个简

单例子。

卡片14.2　台球危机：无限多轨道[14]

一天，我正坐在旧金山机场等飞机，突然想，假如一个台球从虫洞时间机器的两个洞口之间飞过，那么它可能沿两条轨道旅行。一条（a），球无破坏地从两洞口间冲过去；另一条（b），球通过时被撞向右边的洞口，然后进入虫洞，在进洞之前从左洞口出来，与自己相撞，然后飞走。

几个月后，福瓦德（激光干涉仪探测器的先驱者之一（第10章），也是位科幻小说家）发现了满足一切经典物理学定律的第三条轨道，[15]即下面的轨道（C）：碰撞不是发生在两洞口间，而发生在球到达洞口邻近以前。我后来发现，假如球在经历两次碰撞事件之间多次穿过虫洞，那么碰撞还可以越来越早地发生，如（d）和（e）。具体说，在（e）的情形，球沿路线α向上，与它未来的自己碰撞，沿着β进入右洞口，然后穿过虫洞（回到过去），从左洞口出来，又沿γ穿过虫洞（回到更远的过去），然后沿δ再穿过虫洞（回到更更远的过去），从ε出来，去与它自己碰撞，偏向ζ落下。




显然，有无限多条轨道（每一条经过虫洞的次数不同），从完全相同的初始条件（相同的初始位置和速度）出发，都满足经典（而不是量子）的物理学定律。它留给我们的问题是，物理学是不是疯了？或者，我们想知道，物理学定律能用什么办法告诉我们球应该走哪条轨道吗？

时间机器令物理学疯狂了吗？令它失去了对事物演化的预言能力了吗？如果没有，那么物理学定律如何从无限多个可能中选择一条台球会走的轨道呢？

为了寻找答案，克林卡默和我在1989年从经典物理学定律转向了量子定律，为什么呢？因为它们才是我们宇宙最终的法则。

例如，量子引力定律将最终把握引力和时间与空间的结构。爱因斯坦经典的广义相对论引力定律不过是量子引力定律的一种近似——在远离一切奇点，在时空尺度远大于$10^{-33}$厘米时，近似是非常准确的，但毕竟还是一种近似（第13章）。

同样，学生和我用以研究波尔琴斯基怪圈的经典的台球物理学定律，也不过是量子力学定律的一种近似。由于经典定律似乎预言了一些“废话”（无限多个可能的台球轨道），为了更深入地认识，克林卡默和我才转向了量子力学定律。

量子物理学中的“游戏规则”大不同于经典物理学的。在给定的初始条件下，经典定律预言将要发生什么（如台球会走哪条路）；而且，如果没有时间机器，它们的预言是惟一的。量子定律则不同，它们只预言将要发生的事件的概率（例如，球通过空间这个或那个区域的概率），而不是确定性的东西。

从量子力学的这些“游戏规则”看，克林卡默和我根据量子力学定律所得的答案也就不那么令人惊奇了。我们发现，假如球出发以后沿波尔琴斯基怪圈的轨道（图14.9和14.10，在下午t＝3:00时刻），那么它接下去走哪条路线，都有一定的量子力学几率，例如，图14.10（a）的几率为48%，（b）的几率为48%——对无限多经典定律所允许的每一条轨道，它都有一定的（小得多的）概率。任何一次“实验”，球只能走某一条经典路线；但如果我们做大量相同的台球实验，那么其中有48%的球会走轨道（a），有48%走轨道（b），依此类推。

结果多少还是令人满意的。它似乎说明物理学定律可能会很好地使自己适应时间机器。当然也有些令人惊讶的东西，但似乎没有任何怪异的预言，也没有任何解不开的怪圈。[16]实际上，如果《国家调查者》杂志（National Enquirer）听说了，可以很容易打出一个通栏大标题：物理学家证实存在时间机器。（当然，我还是害怕报刊会把这些东西曲解成怪物。）

1988年秋，我们的论文，“虫洞、时间机器和弱能量条件”发表3个月后，《旧金山检查者》（San Francisco Examiner）记者戴维森（Keay Davidson）在《物理学评论通讯》上看到了，于是故事传开

了。

这样一来，事情就更糟了。在那3个月里，至少物理学界很安静，他们在考虑我们的思想，而不是要听什么张扬和吹嘘。

但张扬是挡不住了。物理学家发明时间机器，这是常见的标题。《加利福尼亚》杂志在“发明时间旅行的人”的文章里，甚至登出一张我在帕洛玛山赤膊工作的照片。我很惭愧——不是为照片，而是为那些离谱的宣扬，说我发明了时间机器和时间旅行。[17]事实上，就算物理学定律允许时间机器（在本章最后可以看到，我怀疑这一点），人类现在的技术能力离它还远得很，比洞穴野人离太空旅行还要遥远。

我和两个记者谈过，才发现没有办法抵挡这潮流，也没有办法让他们把故事讲得更准确，还是一个人躲起来吧。我的后勤助理莱昂（Pat Lyon）却被大家包围了，他只好搪塞说：“索恩教授相信，向大家公布研究结果，现在为时尚早。时间机器是否为物理学定律所禁止，等他觉得有了更好的认识后，会为大家写一篇文章来解释的。”

我写这一章，就是在履行那个诺言。

[1]　在多数科幻小说作品中用的是“祖父怪圈”而不是“母子怪圈”。也许，这些小说作家们都是尊重女性的大侠，觉得回到过去谋害一个男子会更心安一些。（原文“matricide paradox ”应为“弑母怪圈”，我回避了“那个”字，觉得这样更好。——译者）

[2]　我们兄弟姐妹四个都很尊敬孝顺母亲，例如，你可以看第7章的那个脚注。我在这儿举的例子是经母亲同意了的。

[3]　3年后，弗里德曼和莫里斯一起设法严格证明了，波通过虫洞回到过去时，确实不会产生解不开的怪圈——只要波线性叠加的方式与卡片10.3讲的相同。[12]

## 良序

1989年2月，大众的喧闹慢慢静下来，埃切维里业、克林卡默和我继续波尔琴斯基怪圈的研究。我飞往蒙大拿波茨曼去演讲，在那儿碰到了米斯纳以前的学生希斯科克（Bill Hiscock）。和看见别的同行一样，我也向他请教他对虫洞和时间机器的看法。我在寻找有力的批评、新颖的思想和独特的观点。

“也许你该研究电磁真空涨落，”希斯科克告诉我，“在无限发达的文明把虫洞变成时间机器时，它们可能会破坏它。”在他的头脑里也有个思想实验：卡洛丽（假定是无限发达的）正带着一个虫洞口坐着我们家的飞船飞回地球，我带着另一个洞口坐在地球上，而虫洞即将成为时间机器（见上面的图14.7和图14.8）。希斯科克在想，电磁真空涨落也可能像图14.8节的辐射那样穿过虫洞，然后与自己碰撞，最后变得无限强烈从而破坏虫洞。

我表示怀疑。一年前，我在从芝加哥回家的路上曾想到，穿过虫洞的辐射不会和自己碰撞产生无限大能量的辐射束，辐射将被分散，从而虫洞不会受到破坏。我相信虫洞也会分散穿过它的电磁真空涨落，从而挽救自己。

另一方面，我想，既然时间机器是那样一个异乎寻常的物理学概念，我们应该考察任何一种可能破坏它的机会。所以，尽管我也怀疑，但还是让我的一个博士后金成旺（Sung-Won Kim）去计算穿过虫洞的真空涨落的行为。

虽然，希斯科克和康科夫斯基（Deborah Konkowski）几年前建立了很好的数学方法和思想，但金和我还是没什么进展，[18]都怨我们自己太笨，没有一个熟悉关于真空涨落的弯曲时空的量子场定律（第13章）。不过，经历了一年的错误以后，我们在1990年2月终于完成

了计算，得到了答案。

答案令我惊讶。尽管虫洞将努力分散真空涨落，但它们似乎会自动再聚集起来（图14.11）。涨落被虫洞分散后，在地球的洞口散开，仿佛到不了飞船；接着，像受到某种神秘力量吸引似的，它们又自动聚向卡洛丽飞船的洞口，通过虫洞回到地球，然后又在洞口散开，又再聚向飞船的洞口，如此反反复复，最后形成一束强大的涨落能量。

这样一束电磁真空涨落有破坏虫洞的能力吗？我们问自己。从1990年2月到9月的8个月，我们一直在同这个问题搏斗。经过几回反复，最后，我们（错误地）认为涨落“大概不会”破坏虫洞。我们自己和几个讨论过结果的同事都觉得论证有力，于是，我们写成一篇文章，交给《物理学评论》。

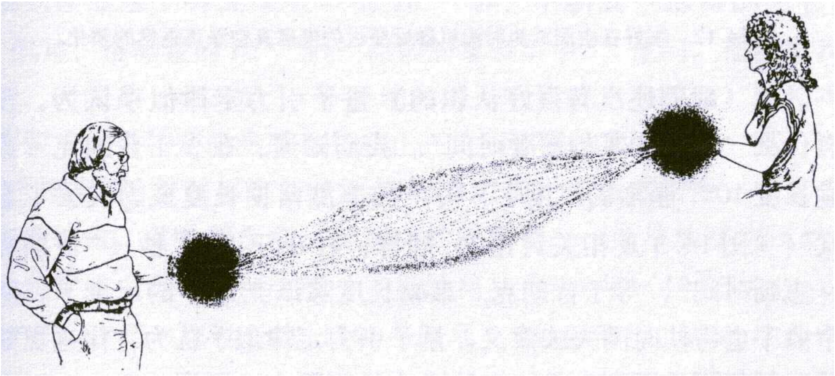


图14.11　当卡洛丽和我在用图14.7的办法努力把虫洞转变为时间机器时，在两个洞口间穿行的电磁真空涨落与自己发生碰撞，产生一束巨大的涨落能量。

我们的论证是这样的：计算表明，在虫洞中往来的电磁真空涨落，只有在近乎为零的短暂时间里才可能无限强大。它们几乎在第一次能用虫洞做时间旅行的瞬间（也就是在虫洞刚成为时间机器时）达到最高峰，然后立刻消失；见图14.12。

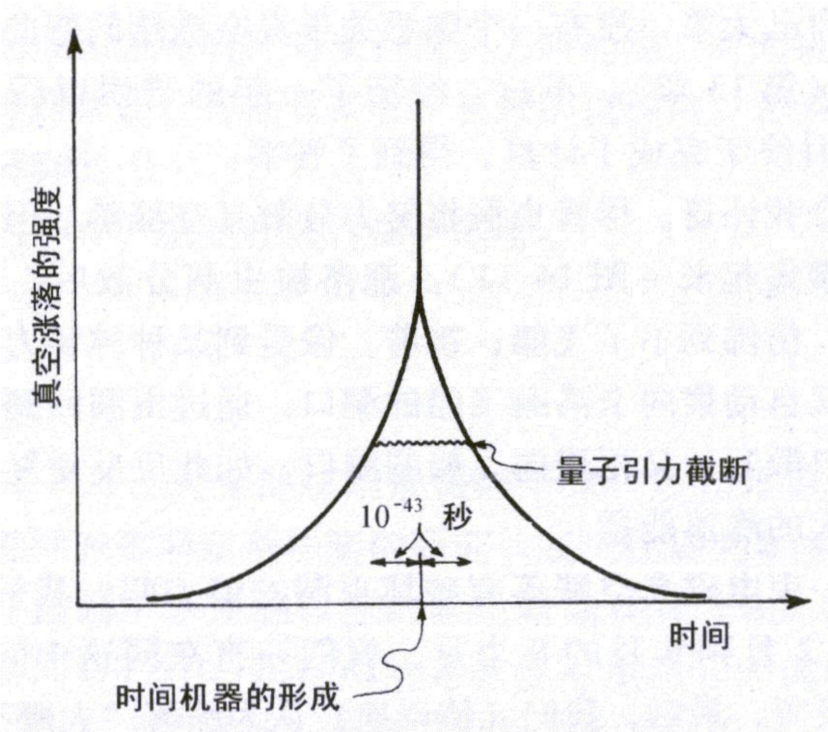



图14.12　刚好在虫洞成为时间机器后穿过的电磁真空涨落强度的演化。

而（我们还没有很好认识的）量子引力定律似乎认为，没有什么“似乎为零的短暂时间”。我们知道，在小于普朗克-惠勒长度$10^{-33}$厘米的尺度下，时空曲率涨落使长度概念失去了意义（见图14.3及相关讨论）；同样，在小于$10^{-43}$秒（“普朗克-惠勒时间”，等于普朗克-惠勒长度除以光速）的尺度下，时空曲率也将使时间失去意义。量子引力定律似乎认为，比这更短的时间间隔是不存在的。在这么小的间隔内，所谓以前、以后和随时间演化的说法都没有意义。

于是，金成旺和我认为，在虫洞间往来的电磁真空涨落一定会停止随时间的演化，也就是，在虫洞成为时间机器的$10^{-43}$秒之前停止增长，量子引力定律一定会中断涨落的生长；而让它只能在时间机器诞生$10^{-43}$秒后再继续生长，那意味着在涨落开始消失以后。在这些时间之间，没有时间，也没有演化（图14.12）。这时，关键的问题是，

在被量子引力中断生长时，往来的涨落有多强？我们的计算确凿无疑：涨落束在停止生长时远远不能破坏虫洞，于是，用我们在文章里的话来说，大概“真空涨落不能阻止类时闭曲线的形成和存在。”（我以前讲过，类时闭曲线就是物理学家说的“时间机器”；“时间机器”在大众中热过一回了，这回我没在文章里用它；不熟悉物理学名词的普通读者，不知道我发表的是关于时间机器的新结果。）

1990年9月，在把文章交给《物理学评论》时，金和我给许多同事寄去了复印件，也给霍金寄了一份。霍金津津有味地读了，不同意。关于真空涨落的计算，他没有什么意见（实际上，弗罗洛夫（Valery Frolov）在莫斯科做的相同计算已经证实了我们的结果[19]）；他反对的是我们对量子引力效应的分析。

霍金同意，量子引力很可能在时间机器产生前$10^{-43}$秒，也就是在涨落变得无限大以前$10^{-43}$秒，中断真空涨落的生长。“但是，谁测量的$10^{-43}$秒？在谁的参照系中？”他问。他提醒我们，时间是“相对的”，不是绝对的，它依赖于参照系。金和我曾假定这个特定的参照系是静止在虫洞咽喉的某个人的。霍金说（大概意思），如果选一个不同的参照系，如涨落自身——或者更准确说，某个随涨落一起运动的观察者——他从地球到飞船，快速穿过虫洞，看到地球-飞船距离从10光年（$10^{19}$厘米）收缩到普朗克-惠勒长度（$10^{-33}$厘米）。霍金猜想，从这个往来的观察者看，量子引力只有在虫洞成为时间机器前$10^{-43}$秒才能决定和中断涨落束的生长。

从静止在虫洞的观察者（金和我依靠的观察者）看，霍金的猜想意味着，量子引力中断涨落生长发生在虫洞成为时间机器$10^{-55}$秒前，而不是$10^{-43}$前——到那个时候，照我们的计算，真空涨落束是足够强大的（但也只不过刚好这么大），可能确实会破坏虫洞。

霍金猜想的量子引力中断的时刻是令人信服的。金和我想了很久，最后认为他很可能是对的。我们想赶在论文发表以前把它改正过

来。[20]

然而，最基本的一点还是不能确定。即使霍金对了，真空涨落束会不会破坏虫洞，仍然远远没有说明——寻找确定的结果，需要我们认识量子引力在时间机器形成那一时刻附近$10^{-95}$秒的间隔内会做些什么。

简单地说，量子引力将虫洞能否成功成为时间机器的答案藏起来了。为了找出答案，我们首先得成为量子引力定律的专家。

霍金对时间机器有着严厉的批评，他认为大自然也憎恶它们；他把这种憎恶表达为一个猜想，一个能维护时间次序的良序猜想，[1]它指出，物理学定律不允许时间机器。[21]（霍金以他特有的幽默，说这个猜想能“保证世界不会破坏历史”。）

霍金猜测，大自然就是通过真空涨落束的生长来加强维护时间顺序的：当我们想做时间机器时，不论用什么样的事物（如虫洞、旋转柱[2]、“宇宙弦”[3]或其他什么东西），在它成为时间机器前，总会有一束真空涨落穿过它，并破坏它。霍金好像已经准备为这个结果下大赌注了。

我不愿成为这个赌局的另一方。我真喜欢同霍金打赌，但我只打获胜机会较大的赌。我本能地感到，如果去赌这个，我准会输的。我与金的计算和弗朗纳根（Eanna Flanagan，我的学生）最近没发表的计算似乎说明霍金很可能是对的。不过，在物理学家深刻认识量子引力定律之前，我们谁也不能肯定。[23]

[1] 原文“Chronology Protection”是“时序保护”，我觉得这在汉语里不像一个“术语”，所以借了一个数学名词，“良序”，前面加“维护时间”的定语，似乎还算恰当。本节小标题也是用的这个词。——译者

[2] 见465页的脚注2。

[3] 普林斯顿大学Richard Gott最近发现，可以通过让两根无限长宇宙弦（一种在宇宙中可能存在也可能不存在的假想物体）以极高速度相对移动来做时间机器。[22]

# 尾声

爱因斯坦遗产的过去和未来，

几个重要角色的今天

爱因斯坦打破牛顿的绝对空间和时间的概念，奠定自己的理论基础，离现在差不多整整一百年了。在这一百年里，爱因斯坦的理论在成长；在他留下的精神财富里，我们看到了时空的弯曲和一堆完全由这弯曲产生的奇异东西：黑洞、引力波、奇点（隐藏的和裸露的）、虫洞和时间机器。

在历史的某些时期里，这些东西都曾被物理学家看成怪物。

·我们在书中看到，爱丁顿、惠勒，甚至爱因斯坦都曾强烈怀疑黑洞；爱丁顿和爱因斯坦没能活着看到他们的错误；而惠勒后来成了黑洞的宣传者。

·40年代和50年代，许多物理学家因为错误相信了他们正在研究的广义相对论的数学解释，曾怀疑引力波（曲率的波动）的存在——不过那该是另一本书的故事，而且怀疑早就没有了。

·奇点是爱因斯坦广义相对论不可避免的结果，这个发现曾极大震撼了大多数物理学家，现在也仍然在震撼着许多人。有人从彭罗斯的宇宙监督猜想（所有奇点都被隐藏着，裸露的奇点是被禁戒的）找到了安慰。然而，不论宇宙监督是否正确，多数物理学家还是习惯了奇点；而且，他们与惠勒一样，期待着未知的量子引力定律来抹平这些奇点——来规定和限制它们的行为，就像牛顿和爱因斯坦的引力定律

规定行星并限定它们绕太阳的轨道一样。

·虫洞和时间机器，即使爱因斯坦广义相对论定律允许它们存在，在今天的大多数物理学家看来也是怪物。不过，我们刚发现，尽管爱因斯坦定律允许虫洞和时间机器的存在，却不能主宰它们的行为；主宰它们的是更严格的弯曲时空的量子场定律和量子引力定律。怀疑的物理学家大概能从这里得到些安慰。假如我们更好认识了那些定律，它们也许会明确地告诉我们，物理学定律总会让宇宙摆脱虫洞和时间机器——或者，也许至少会摆脱时间机器。

在未来的世纪里，在爱因斯坦理论的第二个百年里，我们能等到些什么呢？

我们关于空间、时间和时空弯曲所产生的事物的认识可能会发生革命，它一点儿也不亚于我们在第一个百年里经历过的革命；它的种子已经播下了：

·引力波探测器很快会为我们带来黑洞的观测图像，我们将听到黑洞碰撞的交响曲——充满了弯曲时空在疯狂振荡时的活动信息的交响曲；我们还将从超大规模计算机的模拟中听到它们的回声，领会它们的意义。于是，黑洞成了实验仔细审查的对象。审查结果呢？会令我们惊讶的。

·在未来的百年里，可能很快，用不了多久，某个有远见的物理学家将最终发现并揭开量子引力定律的一切细节。

·有了那些量子引力定律，我们可能会完全了解宇宙时空如何从量子泡沫或大爆炸泡沫中出现，如何存在下去；我们可能知道，那个常被人问到的问题，“大爆炸以前是什么？”有什么意思，还是没有意思；我们可能知道，量子泡沫会不会很容易地产生多个宇宙；时空如何在黑洞中心或大挤压的奇点处遭到毁灭；时空是不是可以再生，如

何再生；我们可能知道，量子引力定律是不是允许（或禁止）时间机器：时间机器总会在它们运行的瞬间自我毁灭吗？

·从牛顿定律到狭义相对论，到广义相对论和量子论，然后到量子引力，这条物理学定律之路并不会终结在量子引力。量子引力定律还将与大自然的其他基本力的定律结合（统一）：电磁力，弱力和强力。也许我们将在未来百年里了解那个统一的细节——同样可能很快，不会等太久，这个统一可能又会从根本上改变我们的宇宙观。然后呢？今天还没有人能预见那统一以外的事情；我相信——而且在你我的有生之年，统一迟早总会到来的。

## 终曲：1993

爱因斯坦最后25年的大部分时间都在徒劳地追寻广义相对论物理学定律与麦克斯韦电磁学定律的统一；他不知道最重要的统一是与量子力学的统一。1955年，76岁的他死在新泽西普林斯顿。

钱德拉塞卡83岁了，还在探寻爱因斯坦场方程的秘密，经常是跟年轻得多的同事们合作。近些年，他教给我们许多关于恒星脉动和引力波碰撞的事情。[1]

茨维基越来越成了实测天体物理学家，而不大是理论家；他还在继续独创一些惹人争论的有远见的思想，不过不是本书的题目。1968年，他离开加州理工学院教授的位置来到瑞士，在那里度过了他追寻自己内心通向真理的道路（“形态学方法”）的余生，1974年去世。

朗道经历了一年（1938～1939）的监狱生活后，虽恢复了智力，却没有了激情；他还是苏联理论物理学家的领袖，也是最严厉的老师。1962年他在车祸中严重受伤，大脑坏了，生活也跟着发生改变，不能再做物理了。他死于1968年，但他的亲密朋友后来说，“对我来说，朗道1962年就死了。”

泽尔多维奇从70年代到80年代一直是世界上最有影响的天体物理学家。不过，1978年，因为人际关系的破裂，他凄凉地离开了他的研究小组（那是世界上有史以来最有力量的一支理论天体物理学家队伍）。他想重建一个年轻人的队伍，但是不很成功；后来，在80年代，他成了全世界天体物理学家和宇宙学家的偶像。1987年，戈尔巴乔夫的政治改革使他第一次能有机会来美国，但没过多久，他就因心脏病在莫斯科去世了。

诺维科夫在泽尔多维奇离开后，成了那个研究群体的领导者。在80年代，他像过去的泽尔多维奇一样，用他的思想火花来激发和团结他的群体。然而，离开了泽尔多维奇，这个小组不过是全世界许多优秀小组中的一个，而不再像以前那样领先了。1991年，苏联解体，接着诺维科夫又做了心脏手术，他觉得活不了多久了，来到丹麦哥本哈根大学，现在在那儿创建新的理论天体物理学中心。

金兹堡77岁了，仍在物理学和天体物理学的几个不同分支里做着前沿研究。1980～1986年，在萨哈洛夫流放高尔基期间，作为他在莫斯科列别德夫研究所的领导，金兹堡没有开除他还保护了他。在戈尔巴乔夫的改革时代，金兹堡和萨哈洛夫都当选为苏联人民下院议员，推行改革。1989年，萨哈洛夫死于心脏病。

奥本海默尽管在1954年的忠诚调查听证会上受到美国政府批判，但在多数物理学家心目中他还是英雄。他以后没有再做过研究，但仍和大多数物理学分支保持着密切联系；他扶持了许多年轻的物理学家，他们都愿意跟他讨论物理学问题，发展自己的思想。1967年，他死于癌症。

惠勒82岁，继续追寻量子力学与广义相对论的结合，继续用他的演说和著作来激发年轻的一代，最近（1990）他有一本特别值得注意的书：《引力和时空之旅》。

彭罗斯跟惠勒和许多其他人一样，对广义相对论和子力学的结合着迷了——而且未来的量子引力定律可能会从这个结合中产生。他在一本为非物理学家读者写的书（《皇帝新脑》[2]，1989）里描绘了自己的非传统思想。很多物理学家怀疑他的观点，但他从来都是这样的，而且我们也看到了他以前好多次都对了……

霍金也一样继续为量子引力定律着迷，而他最感兴趣的问题是，那些定律关于宇宙起源会预言些什么。跟彭罗斯一样，他也为非物理学家写了一本书（《时间简史》，1988），谈自己的思想。尽管患有肌萎缩性脊髓侧索硬化（ALS），他还是活得很结实。

[1] 两年后（1995）的8月21日，这位1983年度的诺贝尔奖获得者在芝加哥逝世。——译者

[2] 这本书和下面霍金那本的中译本都在我们这个“第一推动丛书”的系列里。——译者

# 致谢

感谢曾影响过这本书的

朋友们和同事们

这本书是在爱莱茵（Elaine Hawkes Watson）对宇宙的无限好奇心的激发下开始写的。在15年断断续续的写作中，我得到了我的家庭和几个亲密朋友的巨大鼓励和支持：琳达、卡丽丝、布雷特、阿里森、格里戈利（Estelle Gregory）、舒梅克（Bonnie Schumaker），特别是我的妻子，卡洛丽·温斯顿（Carolee Win-stein）。

我要感谢许多物理学家、天文学家、天体物理学家同事，他们答应了我的录音采访，向我讲述了书中的大量历史事件和研究经过。他们的名字列在参考文献的开头。

我的4位同事，布拉金斯基、霍金、伊斯雷尔和萨根读了全部书稿，提出了具体的批评。很多人还读过一章或几章，为我澄清了许多重要的历史事件和科学细节。他们是：Vladimir Belinsky,Roger Blandford,Carlton Caves,S.Chandrasekhar,Ronald Drever,Vitaly Ginzburg,Jesse Greenstein,Isaac Khalatnikov,Igor Novikov,Roger Penrose,Dennis Sciama,Robert Serber,Robert Spero,Alexi Starobinsky,Rochus Vogt,Robert Wald,John Wheeler和Yakov Borisovich Zel'dovich。没有他们的帮助，这本书不会像现在这样准确。不过，读者应该想到，我的同事不会完全赞同我对我们故事的解释，难免会有不同的观点。在书中，为了让读者更容易理解，我遵循了自己的观点（通常受过同事批评的重要影响，但不是全部）。为尊

重历史，我在注释中暴露了某些矛盾。

Lynda Obst严厉批评了第一稿的许多章节，我谢谢她；K.C.Cole批评了第二稿，还耐心地一篇一篇地为我提出建议，最后才修改成现在这样。我要特别感谢K.C.。我还要谢谢De-bra Makay,她是比我更彻底的理想主义者，一丝不苟地检查了最后的手稿。

本书的提高还大大得益于几位非物理学家读者的批评：Ludmila（Lily）Birladeanu,Doris Drücker,Linda Feferman,Rebecca Lewthwaite,Peter Lyman,Deanna Metzger,Phil Richman,Barrie Thorne,Alison Thorne,还有卡洛丽。我谢谢他们。我还要感谢Helen Knudsen为我找到了好多参考文献和事实——有些困难是人们想象不到的。

我幸运地在Heiz Pagel的《宇宙密码》中偶然看到Matthew Zimet的令人赏心悦目的画，也请他为这本书画了插图，真是增色不少。

最后，感谢联邦基金会图书计划，特别感谢在我完成这本书的几年中耐心支持和信任我的Alexander G.Bearn和Antonina W.Bouis——以及W.W.Norton公司和该公司的Ed Barber。

# 人物

在本书不同地方多次出现过的人物

说明：下面的叙述只是为了回忆每个人在书中哪些地方出现过，便于前后参照，不能作为个人的传略。（多数人在科学的其他领域有过重要贡献，与本书无关，就没有在此列举了。）列在这里的人物，不是看他的贡献大小，而是因为他们在书中的不同地方出现过多次。

巴德（Baade,Walter,1893～1960）生于德国，美籍光学天文学家，与茨维基提出超新星概念及其与中子星的联系（5）；[1]确认与宇宙射电源相关的星系（9）。

巴丁（Bardeen,James Maxwell，1939～）美国理论物理学家，证明宇宙中多数黑洞在快速旋转；与彼德森预言黑洞旋转对周围吸积盘的影响（9）；与卡特尔和霍金发现黑洞力学四定律（黑洞演化定律）（12）。

贝肯斯坦（Bekenstein,Jacob，1947～）以色列理论物理学家，惠勒的学生。与哈特尔证明不能通过黑洞外面的任何研究判别形成黑洞的材料的粒子类型（7）；提出黑洞表面积相当于它的熵，与霍金就此争论，最后获胜（12）。

玻尔（Bohr,Niels Hendrik David,1885～1962）丹麦理论物理学家，诺贝尔奖获得者[1922]，量子力学创始人之一，20世纪中叶许多大物理学家（包括朗道和惠勒）的导师；曾指导钱德拉塞卡与爱丁顿的论战（4）；试图救朗道（5）；与惠勒提出核裂变理论（6）。

布拉金斯基（Braginsky,Vladimir Borisovich,1931～）俄罗斯实验物理学家。发现物理学测量精度（包括引力波探测器的精度）的量子力学极限（10）；发明克服量子极限的“量子无破坏”装置（10）。

卡特尔（Carter,Brandon，1942～）澳大利亚理论物理学家，英国剑桥席艾玛的学生，后移居法国。阐明旋转黑洞性质（7）；与人证明黑洞无毛（7）；与巴丁和霍金发现黑洞力学四定律（黑洞演化定律）（12）。

钱德拉塞卡（Chandrasekhar,Subrahmanyan,1910～[1995]）生于印度，美籍天体物理学家，诺贝尔奖获得者[1983]；证明白矮星存在极大质量，就预言正确性与爱丁顿争论（4）；发展黑洞微扰理论（7）。

爱丁顿（Eddington,Arthur Stanley，1882～1944）英国天体物理学家，爱因斯坦广义相对论定律的早期倡导者（3）；黑洞概念和钱德拉塞卡白矮星极限质量的强烈反对者（3,4）。

爱因斯坦（Einstein,Albert，1879～1955）生于德国，瑞士/美国理论物理学家，诺贝尔奖获得者[1921]；创立狭义相对论（1）和广义相对论（2）；证明光同时既是粒子也是波（4）；反对黑洞概念（3）。

格罗赫（Geroch,Robert,1942～）美国理论物理学家，惠勒的学生；与人发展黑洞分析的整体方法（13）；证明空间拓扑只有在时间机器的产生过程中才会发生改变（如形成虫洞）（14）；与瓦尔德第一次提出时间机器可能在形成时毁灭（14）。

贾柯尼（Giacconi,Riccardo,1931～）生于意大利，美国实验物理学家和天体物理学家；1962年领导一个小组利用火箭探测器首次发现X射线星（8）；设计并建造“自由”号X射线卫星，第一次发现天鹅

X-1是黑洞的强X射线证据（8）。

金兹堡（Ginzburg,Vitaly Lazarevich，1916～）苏联理论物理学家；发明苏联氢弹的LiD燃料，然后被开除出氢弹计划（6）；与朗道解释超导性起源（6，9）；发现第一个黑洞无毛证据（7）；提出宇宙射电波的同步辐射起源（9）。

格林斯坦（Greenstein,Jesse,L.,1909～）美国光学天文学家，茨维基的同事（5）；与惠普发现不可能[用传统观点]解释宇宙射电波（9）；激发了美国射电天文学研究的开端（9）；与施米特发现类星体（9）。

哈特尔（Hartle,James　B.,1939～）惠勒的学生；与贝肯斯坦证明不能通过黑洞外面的任何研究判别形成黑洞的材料的粒子类型（7）；与霍金发现黑洞视界演化定律（12）；与霍金正在探索量子引力定律（13）。

霍金（Hawking,Stephen　W.,1942～）英国理论物理学家，席艾玛的学生；完成了黑洞无毛证明的关键部分（7）；与巴丁和哈特尔发现黑洞力学四定律（黑洞演化定律）（12）；发现在忽略量子力学定律条件下黑洞表面积只能增加，在量子力学条件下黑洞会蒸发和收缩（12）；证明小黑洞能在大爆炸时产生；基于天文学家未发现黑洞蒸发所产生的γ射线的事实，与帕奇提出原生小黑洞的观测极限（12）；发展黑洞分析的整体（拓扑学）方法（13）；与彭罗斯证明大爆炸包含奇点（13）；提出“良序猜想”，认为通过真空涨落在时间机器产生时将其破坏而保持时序（14）；与索恩就天鹅X-1是否为黑洞（8）和裸奇点能否在宇宙中形成打赌（13）。

伊斯雷尔（Israel,Werner,1931～）生于南非，加拿大理论物理学家；证明每个非旋转黑洞一定是球形的，提出黑洞通过辐射“脱毛”的证据（7）.；发现黑洞表面积只能增加，但未认识这一结果的意

义（12）；与彭罗斯和奥里证明黑洞奇点周围的潮汐力随黑洞年龄而减弱（13）；探索黑洞研究早期历史（3）。

克尔（Kerr,Roy,P.,1934～）新西兰数学家；发现爱因斯坦场方程的旋转黑洞解："克尔解"（7）。

朗道（Landau,Lev Davidovich,1908～1968）苏联理论物理学家，诺贝尔奖获得者[1962]；30年代将西方理论物理学带回苏联（5，13）；将星体热量解释为星体物质被中心中子核所捕获的结果，从而激发奥本海默对中子星和黑洞的研究（5）；在斯大林大恐怖时期被捕，获释后提出超流体理论（5）；投身苏联核武器研究（6）。

拉普拉斯（Laplace,Pierre Simon，1749～1827）法国自然哲学家；提出并普及牛顿物理学定律下的暗星（黑洞）概念（3，6）。

洛伦兹（Lorentz,Hendrik Antoon,1853～1928）荷兰理论物理学家，诺贝尔奖获得者[1902]；为狭义相对论定律奠定基础，最重要的是洛伦兹-费兹杰拉德收缩和时间膨胀（1）；爱因斯坦创立广义相对论定律时的朋友和伙伴（2）。

麦克斯韦（Maxwell,James Clerk，1831～1879）英国理论物理学家，发展了电磁学定律（1）。

米歇尔（Michell,John,1724～1793）英国自然哲学家；提出并普及牛顿物理学定律下的暗星（黑洞）概念（3，6）。

迈克尔逊（Michelson,Albert Abraham,1852～1931）生于德国的美籍实验物理学家，诺贝尔奖获得者[1907]；发明干涉仪测量技术（1）；以那些技术发现光速独立于观测者在宇宙中的运动速度（1）。

闵可夫斯基（Minkowski,Hermann，1864～1909）德国理论物理学家，爱因斯坦的老师（1）；将空间和时间统一为时空（2）。

米斯纳（Misner,Charles W.,1932～）美国理论物理学家，惠勒的学生；发明用嵌入图描绘坍缩恒星如何形成黑洞（6）；组织对黑洞研究“黄金年代”有重要贡献的一个研究小组（7）；发现在旋转黑洞附近传播的电磁波和其他波能从黑洞汲取旋转能并用来放大自己（12）；发现奇点附近潮汐引力的“搅拌”振荡（13）。

牛顿（Newton,Isaac,1642-1727）英国自然哲学家，创立牛顿物理学与绝对空间和绝对时间概念的基础（1）；创立牛顿引力定律（2）。

诺维科夫（Novikov,Igor Dmitrievich，1935～）苏联理论物理学家和天体物理学家，泽尔多维奇的学生；与多罗什科维奇和泽尔多维奇发现黑洞无毛的某些重要原始证据（7）；与泽尔多维奇提出在银河系寻找黑洞的天文学方法，似乎最终获得成功（8）；与索恩提出黑洞周围的吸积结构理论（12）；与多罗什科维奇预言黑洞内部潮汐力随黑洞年龄而改变（13）；研究物理学定律是否允许时间机器（14）。

奥本海默（Oppenheimer,J.Robert,1904～1967）美国理论物理学家；30年代将理论物理从西欧带回美国（5）；与塞伯否定朗道关于恒星可能由中子核保持热量的结论，与沃尔科夫证明存在中子星极大质量（5）；与斯尼德用高度理想化模型说明大质量恒星在死亡时会坍缩形成黑洞并阐明坍缩的某些重要特征（6）；领导美国原子弹计划，一开始就反对氢弹计划，后来赞同，因此被认为对国家不忠（6）；与惠勒争论坍缩是否形成黑洞（6）。

彭罗斯（Penrose,Roger,1931～）英国数学家和理论物理学家，席艾玛的门生；猜测黑洞通过辐射失去毛（7）；发现旋转黑洞在视

界外空间旋涡里贮藏大量可以汲取的能量（7）；提出黑洞显视界的概念（12，13）；发现黑洞表面积只能增大，但未认识这一结论的重要意义（12）；创立并发展黑洞分析的整体（拓扑学）方法（13）；证明黑洞一定含有奇点，与霍金证明大爆炸含有奇点（13）；提出物理学定律严禁在宇宙中形成裸奇点的宇宙监督猜想（13）。

普雷斯（Press,William　H.,1948～）美国理论物理和天体物理学家，索恩的学生；与特奥科尔斯基证明黑洞关于小扰动稳定（7，12）；发现黑洞脉动（7）；亲历黑洞研究黄金年代的结束（7）。

普赖斯（Price,Richard　H.,1943～）美国理论物理和天体物理学家，索恩的学生；确定地证明黑洞通过辐射失去毛，证明能被辐射的东西将被完全辐射（7）；看到黑洞脉动的证据，但未认识其意义（7）；与人提出黑洞的膜规范（11）；为索恩的时间机器研究担心（14）。

里斯（Rees,Martin,1942～）英国天体物理学家，席艾玛的学生；提出模型解释黑洞从其伴星吸积气体的双星系特征（8）；提出射电星系巨射电叶能源来自通过星系核心的能量束，与布兰福德发展能量束具体模型（9）；与布兰福德等人用模型解释超大质量黑洞如何为射电星系、类星体和活动星系核提供能量（9）。

萨哈洛夫（Sakharov,Andrei　Dmitrievich,1921～1989）苏联理论物理学家；为苏联氢弹奠定重要思想基础（6）；泽尔多维奇的亲密朋友、伙伴和竞争对手（6，7）；后来成为著名异议人士，苏联开放后成为英雄。

史瓦西（Schwaraschild,Karl,1876～1916）德国天体物理学家，发现爱因斯坦场方程的史瓦西解，描绘了非旋转静态或坍缩恒星的时空几何，也描绘了非旋转黑洞（3）；发现爱因斯坦场方程在常密度星体内部的解——爱因斯坦曾以此论证黑洞不能存在（3）。

席艾玛（Sciama,Dennis,1926～）英国天体物理学家，英国黑洞研究者的导师（7,13）。

特奥科尔斯基（Teukolsky,Saul A.,1947～）生于南非，美国理论物理学家，索恩的学生；提出并发展旋转黑洞的微扰分析方法，与普雷斯用此方法证明黑洞相对于微扰是稳定的（7，12）；与夏皮罗发现物理学定律可能允许在宇宙中形成裸奇点的证据（13）。

索恩（Thorne,Kip,1940～）美国理论物理学家，惠勒的学生；提出黑洞能在坍缩恒星内形成的环猜想，并为它找到证据（7）；根据天体物理学资料估计引力波并提出探测引力波的思想和计划（10）；与人发展黑洞的膜规范（11）；提出黑洞熵的统计学起源（12）；通过虫洞和时间机器的思想实验考察物理学定律（14）。

瓦尔德（Wald,Robert M.,1947～）美国理论物理学家，惠勒的学生；发展并应用特奥科尔斯基的黑洞微扰分析方法（7）；与人发现电场在黑洞外的行为特征——成为膜规范的基础（11）；发展黑洞蒸发理论并用于黑洞熵的起源（12）；与格罗赫首次提出时间机器可能在形成时被毁灭（14）。

韦伯（Weber,Joseph,1919～）美国实验物理学家；发明世界第一台引力波探测器（“棒探测器”）并参与发明引力波干涉仪探测器（10）；公认的引力波探测之父。

惠勒（Wheeler,John Archibald,1911～）美国理论物理学家，美国黑洞与广义相对论其他方面众多研究者的导师（7）；与哈里森和若野提出冷死物质的状态方程和完整的冷死星编目表，从而巩固了大质量恒星死后一定形成黑洞的证据（5）；

与玻尔提出核裂变理论（6）；领导设计美国第一颗氢弹的小组（6）；与奥本海默争论黑洞不能形成，后来成为黑洞的主要拥护者

（6）；发明“黑洞”（6）和“黑洞无毛”（7）的名词；论证引力坍缩恒星的“最终状态问题”是认识广义相对论与量子力学结合的关键，在论证中预言了霍金黑洞能蒸发的发现（6，13）；发展量子引力的基础，最重要的是设想并发展量子泡沫的概念，现在我们猜测它是构成奇点的基元（13）；提出普朗克-惠勒长度和面积（12，13，14）。

泽尔多维奇（Zel'dovich,Yakov Borisovich,1914～1987）苏联理论物理学家和天体物理学家，苏联天体物理学家的导师（7）；发展核的链式反应理论（5）：提出苏联原子弹和氢弹基础的关键思想，并领导一个原子弹设计小组（6）；与多罗什科维奇和诺维科夫早就发现黑洞无毛的证据（7）；提出几个寻找黑洞的天文学方法，其中之一似乎最后成功了（8）；独立于萨尔皮特提出超大质量黑洞是类星体和射电星系的能源（9）；猜想并与斯塔罗宾斯基证明，量子力学定律可能导致旋转黑洞辐射而失去旋转，但后来却反对霍金关于非旋转黑洞能辐射和蒸发的证明（12）。

茨维基（Zwicky,Fritz,1898～1974）生于瑞士，美籍理论物理学家、天体物理学家和光学天文学家；与巴德认定超新星是一类天体并提出它们的能源来自正常星变成中子星时所释放的能量（5）。

[1] 括号中的数字代表第几章。

# 年表

## 关于事件、观点和发现的年表

1687　牛顿发表《原理》，[1]建立绝对空间和时间概念、运动定律和引力定律。[1]

1783，1795　米歇尔和拉普拉斯用牛顿的运动、引力和光的定律提出牛顿黑洞的概念。[3]

1864　麦克斯韦建立统一的电磁学定律。[1]

1887　迈克尔逊和莫雷以实验证明光速独立于地球在绝对空间中运动的速度。[1]

1905　爱因斯坦证明空间和时间是相对的，不是绝对的，创立狭义相对论。[1]

爱因斯坦证明电磁波在某些条件下表现粒子行为，从而启发了量子力学基础的波粒二象性概念。[4]

1907　爱因斯坦迈出广义相对论第一步，建立局部惯性系和等效原理，导出时间的引力膨胀。[2]

1908　闵可夫斯基将空间和时间统一为绝对的四维时空。[2]

1912　爱因斯坦发现时空是弯曲的，潮汐引力是曲率的外在表现。[2]

1915　爱因斯坦和希尔伯特独立建立爱因斯坦场方程（描写物质如何扭曲时空），从而完成广义相对论。[2]

1916　史瓦西发现爱因斯坦场方程的史瓦西解，后来证明它描写了非旋转无电荷的黑洞。[3]

弗拉姆发现，适当选择拓扑，史瓦西解能描写虫洞。[14]

1916，1918　雷斯纳和诺德斯特勒姆发现后来用以描写无旋转带电黑洞的爱因斯坦场方程解。[7]

1926　爱丁顿诘难白矮星，攻击黑洞的实在性。[4]

薛定谔和海森伯在别人工作基础上完成量子力学的建立。[4]

福勒用量子力学定律发现白矮星之谜在于电子简并。[4]

1930　钱德拉塞卡发现存在白矮星的极大质量。[4]

1932　查德威克发现中子。[5]

央斯基发现宇宙射电波。[9]

1933　朗道在苏联建立研究小组，传播西方理论物理。[5，13]

巴德和茨维基认证超新星，提出中子星概念，说明超新星能源来自星核形成中子星时的坍缩。[5]

1935　钱德拉塞卡完善白矮星极大质量的证明，爱丁顿批评他的研究。[4]

1935～1939　苏联大恐怖。[5，6]

1937　格林斯坦和惠普说明不能用已知天体物理学过程解释央斯基的宇宙射电波。[9]

朗道为逃脱入狱和死亡提出恒星热量的能源来自流向中心核的物质。[5]

1938　朗道以德国间谍罪在莫斯科入狱。[5]

奥本海默和塞伯否定朗道的恒星中子核热源；奥本海默和沃尔科夫证明中子星存在最大质量。[5]

贝特和克里奇菲尔德证明太阳和恒星的热量来自核燃烧。[5]

1939　濒临死亡的朗道出狱。[5]

爱因斯坦论证黑洞不能在真实宇宙中存在。[4]

奥本海默和斯尼德通过高度理想化计算证明坍缩恒星形成黑洞，（疑惑地）发现，在外面看来，坍缩在视界冻结，而从恒星表面看，并非如此。[6]

雷伯发现来自遥远星系的宇宙射电波，但他不知道所看到的是什么。[9]

玻尔和惠勒提出核裂变理论。[6]

哈里顿和泽尔多维奇提出核裂变的链式反应理论。[6]

德军侵占波兰，第二次世界大战爆发。

1942　美国在奥本海默领导下启动原子弹紧急计划。[6]

1943　苏联开始设计核反应堆和原子弹，水平较低；泽尔多维奇

是主要理论家。[6]

1945　美国在广岛和长崎投放原子弹，第二次世界大战结束。

苏联低水平超弹计划开始启动。[6]

苏联启动原子弹紧急计划，泽尔多维奇是主要理论家。[6]

1946　弗里德曼和他的小组用缴获的德国V-2火箭发射第一台地球大气上空的天文学仪器。[8]

英国和澳大利亚实验物理学家开始建造射电望远镜和射电干涉仪。[9]

1948　泽尔多维奇、萨哈洛夫、金兹堡等人在苏联开始设计超弹（氢弹）；金兹堡发明LiD燃料，萨哈洛夫提出“千层饼”设计。[6]

1949　苏联爆炸第一颗原子弹，平息了美国关于超弹紧急计划的争论。苏联直接实施超弹计划。[6]

1950　美国启动超弹紧急计划。[6]

凯本海尔和金兹堡认识到宇宙射电波是宇宙线电子在星际磁场中的涡旋运动产生的。[9]

亚历山大洛夫和皮苗诺夫将拓扑学工具引进弯曲时空的数学研究，但是结果不好。[13]

1951　特勒和乌拉姆在美国提出威力可以任意大的“真”超弹思想；惠勒在此基础上组队设计并在计算机上模拟。[6]

史密斯为巴德提供1弧分误差区间的天鹅A射电源，巴德用光学望远镜发现天鹅A为一遥远星系——一个“射电星系”。[9]

1952　美国爆炸第一个氢弹装置，它质量太大，不可能以飞机或火箭装载，不过用的是特勒-乌拉姆的思想，并以惠勒小组的设计工作为基础。[6]

1953　惠勒开始研究广义相对论。[6]

詹尼森和古普塔发现星系的射电波是它相对的两片巨叶产生的。[9]

斯大林去世。[6]

苏联在金兹堡和萨哈洛夫思想基础上爆炸第一颗氢弹。由于它的能量不能任意大，苏联科学家宣称它不是“真”超弹。[6]

1954　萨哈洛夫和泽尔多维奇提出“真”超弹的特勒-乌拉姆思想。[6]

美国在特勒-乌拉姆和萨哈洛夫-泽尔多维奇思想基础上爆炸第一颗真正的超弹。[6]

特勒在奥本海默忠贞调查听证会上提出不利证词，认为奥本海默危害国家安全。[6]

1955　苏联在特勒-乌拉姆和萨哈洛夫-泽尔多维奇思想基础上爆炸第一颗真正的超弹。[6]

惠勒提出引力真空涨落概念，认为涨落从普朗克-惠勒长度的尺度长大，在此尺度下时空被量子泡沫所取代。[12，13，14]

1957　惠勒、哈里森和若野提出冷死物质概念，为所有可能的冷死星编目。他们的编目充实了大质量恒星在死亡时必然坍缩成黑洞的结论。[5]

惠勒小组研究虫洞；里奇和惠勒提出虫洞小扰动的微扰分析方法；后来被用于黑洞微扰的研究。[7，14]

惠勒提出星体坍缩的最终状态是研究的最终目标；反对奥本海默的最终状态必然藏在黑洞中的思想。[6，13]

1958　芬克尔斯坦发现史瓦西几何的新参照系，解决了1939年的奥本海默-斯尼德悖论：为什么从外面看，坍缩恒星在临界周长冻结，而从里面看，坍缩会经过临界周长？[6]

1958～1960　惠勒逐渐接受黑洞概念，成为主要拥护者。[6]

1959　惠勒认为大挤压或黑洞中形成的时空奇点由量子引力决定，可能由量子泡沫构成。[13]

布尔比奇证明射电星系巨叶所包含的磁能和动能相当于1千万个太阳质量完全转化为纯能量。[9]

1960　韦伯开始建造引力波的棒探测器。[10]

克鲁斯卡证明，如果球星虫洞没有任何事物穿过，它将很快湮灭而不能通行。[14]

格瑞弗斯和布雷尔发现，爱因斯坦场方程的雷斯纳-诺德斯特勒姆解描写了球形荷电黑洞和虫洞。[7]他们的研究（错误地）认为，不可能从我们宇宙的黑洞内部通过超空间进入某个别的宇宙。[13]

1961　卡拉特尼科夫和栗弗席兹（错误地）论证爱因斯坦场方程不允许有随机形变曲率的奇点存在，从而认为在真实黑洞和宇宙大收缩中不能形成奇点。[13]

1961～1962　泽尔多维奇开始研究天体物理学和广义相对论，

召集诺维科夫等人，建立自己的研究队伍。[6]

1962　索恩开始在惠勒指导下做研究，启发了后来的环猜想。[7]

贾柯尼和他的小组用装在地球上空探测火箭上的盖革计数器发现宇宙X射线。[8]

1963　克尔发现爱因斯坦场方程解。[7]

施米特、格林斯坦和桑达奇发现类星体。[9]

1964　黑洞理论研究的黄金时代开始。[7]

彭罗斯在相对论研究中引入拓扑学工具，证明所有黑洞内部都必然存在着奇点。[13]

金兹堡以及多罗什克维奇、诺维科夫和泽尔多维奇发现黑洞无毛的第一个证据。[7]

美国的科尔盖特、麦和怀特，苏联的波杜利兹、伊姆舍尼克和纳杰任以原子弹设计的计算机语言来模拟实际的星体核坍缩；他们证实了1934年茨维基关于小质量坍缩形成中子星并触发超新星的猜想，也证实了1939年奥本海默-斯尼德关于大质量坍缩形成黑洞的结果。[6]

泽尔多维奇、古塞诺夫和萨尔皮特第一次提出在现实宇宙中寻找黑洞的方法。[8]

萨尔皮特和泽尔多维奇（正确地）猜想类星体和射电星系的能量来自超大质量黑洞。[9]

弗里德曼和他的小组用装在火箭上的盖革计数器发现天鹅X-1。[8]

1965　波耶以及林凯斯特、卡特和彭罗斯发现爱因斯坦场方程的克尔解描写了旋转黑洞。[7]

1966　泽尔多维奇和诺维科夫提出在X射线星和光学星构成的双星系中寻找黑洞，70年代获得成功（也许是的）。[8]

格罗赫证明只有当时间机器出现（或至少在瞬间出现）时，空间拓扑才可能发生非量子力学的改变。[14]

1967　惠勒为黑洞命名。[6]

伊斯雷尔严格证明第一个黑洞无毛猜想：非旋转黑洞一定是完全球形的。[7]

1968　彭罗斯论证不可能从我们宇宙的黑洞内部通过超空间进入别的宇宙；70年代将有人证明他的论证是对的。[13]

卡特尔发现旋转黑洞周围空间旋涡的性质和它对下落粒子的影响。[7]

米斯纳与别林斯基、卡拉特尼科夫和栗弗席兹独立发现爱因斯坦场方程的“搅拌”振荡的奇点。[13]

1969　霍金和彭罗斯证明宇宙一定在大爆炸膨胀之初有过奇点。[13]

别林斯基（B）、卡拉特尼科夫（K）和栗弗席兹（L）发现爱因斯坦场方程的振荡的BKL奇点解；他们证明它的弯曲时空有随机形变，从而认为是在黑洞内部和大挤压时形成的一类奇点。[13]

彭罗斯发现旋转黑洞在周围空间的旋涡运动中贮藏着能量，这样的旋转能可以利用。[7]

彭罗斯提出物理定律严禁形成裸奇点的宇宙监督猜想。[13]

林登-贝尔提出巨黑洞存在于星系核并被吸积盘所包围。[9]

克里斯托多罗发现黑洞缓慢吸积物质时的演化与热力学定律间的相似性。[12]

韦伯宣布引力波存在的证据（但那只是暂时的），激发许多实验家开始建造棒探测器，1975年会明白韦伯没有看到引力波。[10]

布拉金斯基发现引力波探测器的灵敏度存在一种量子力学极限。[10]

1970　巴丁证明气体的吸积可能使宇宙中的典型黑洞很快地旋转。[9]

普赖斯在彭罗斯、诺维科夫以及切斯、德拉克鲁兹和伊斯雷尔工作的基础上证明黑洞通过辐射“脱毛”，他还证明任何可能被辐射的东西都将辐射殆尽。[7]

霍金提出黑洞绝对视界的概念，证明其表面积总在增加。[12]

贾柯尼小组建成第一个卫星X射线探测器“自由”，并发射进入轨道。[8]

1971　X射线、射电波和光学观测证据的结合证明天鹅X-1是环绕一颗正常恒星的黑洞。[8]

MIT的外斯和休斯公司的福瓦德率先用干涉仪探测引力波。[10]

里斯提出射电星系的巨射电叶的能量来自星系核射出的喷流。[9]

汉尼和鲁菲尼提出黑洞表面电荷概念，成为膜规范的基础。[11]

普雷斯发现黑洞会脉动。[7]

泽尔多维奇猜测旋转黑洞会辐射，并与斯塔罗宾斯基用弯曲时空的量子场定律证明这个猜想。[12]

霍金提出“原生”小黑洞可能是在大爆炸中形成的。[12]

1972　卡特尔在霍金和伊斯雷尔研究的基础上证明无电荷旋转黑洞的无毛猜想（后来罗宾逊补充了一些技术细节）。他说明这类黑洞总可以用爱因斯坦方程的克尔解来描写。[7]

索恩提出环猜想作为黑洞能否形成的判据。[7]

贝肯斯坦猜测黑洞的表面积表现为熵，而黑洞的熵是黑洞形成方式总数的对数。霍金坚决反对这一猜想。[12]

巴丁、卡特尔和霍金建立在形式上与热力学定律完全相同的黑洞演化定律，但仍然坚持视界表面积不可能表现为黑洞的熵。[12]

特奥科尔斯基发展描写旋转黑洞脉动的微扰方法。[7]

1973　普雷斯和特奥科尔斯基证明旋转黑洞的脉动是稳定的，不会因补充黑洞的旋转能而增强。[7]

1974　霍金证明一切旋转和不旋转的黑洞都表现出一个正比于表面引力的温度并因此而辐射，从而蒸发。然后，他改变了主张，赞成黑洞力学定律就是黑洞的热力学定律；他也同意了贝肯斯坦的猜想，黑洞表面积就是黑洞的熵。[12]

1974～1978　布兰福德、里斯和林登-贝尔提出星系核和类星体中超大质量黑洞产生喷流的几种方式。[9]

1975　巴丁和彼德森证明旋转黑洞周围的空间旋涡能像陀螺一样保持喷流方向。[9]

钱德拉塞卡为了建立完整的黑洞微扰的数学描述开始他5年的探索。[7]

昂鲁什和戴维斯推测，在黑洞视界上方的加速观测者会看到黑洞被热粒子大气所包围，这些粒子的逃逸即说明了黑洞的蒸发。[12]

帕奇计算了黑洞辐射的粒子谱。霍金和帕奇通过宇宙$\gamma$射线的观测数据推测，每立方光年的空间中正在蒸发的原生小黑洞不会超过300个。[12]

一群年轻的研究者宣告黑洞理论研究的黄金年代结束。[7]

1977　吉本斯和霍金证明贝肯斯坦的猜想：黑洞的熵是其可能形成方式总数的对数。[12]

射电天文学家用干涉仪发现从星系中心黑洞汲取能量输入巨射电叶的喷流。[9]

布兰福德和茨纳耶克证明，穿过黑洞视界的磁场能汲取黑洞的旋转能，并为类星体和射电星系提供能源。[9]

茨纳耶克和达莫尔提出黑洞视界的膜描述方法。[11]

布拉金斯基和同事，凯维斯、索恩和同事为克服引力波探测棒的量子极限设计量子无破坏传感器。[10]

1978　贾柯尼小组完成第一台高分辨率X射线望远镜“爱因斯坦”的建造，并发射送入轨道。[8]

1979　汤尼斯等人发现银河系存在300万个太阳质量黑洞的证

据。[9]

德雷维尔在加州理工学院启动干涉仪引力波探测计划。[10]

1982　邦汀和马祖尔证明带电旋转黑洞的无毛猜想。[7]

1983～1988　芬尼等人在黑洞基础上建立综合模型解释类星体和射电星系所有细节。[9]

1984　国家科学基金会强令加州理工学院和麻省理工学院的两个引力波探测计划合并，促成LIGO计划。[10]

雷德蒙特（在伊尔德莱研究基础上）证明落进球状空虫洞的辐射将达到高能并加速虫洞的湮灭。[14]

1985～1993　索恩、莫里斯、尤尔泽维尔、弗里德曼、诺维科夫等人提出物理学定律是否允许虫洞旅行和时间机器问题。[14]

1987　伏格特领导LIGO计划，开始大步向前。[10]

1990　金成旺和索恩证明，不论用什么方法做时间机器，总会有一束强真空涨落在它产生的瞬间穿过去。[14]

1991　霍金提出维护时序的良序猜想（物理学定律严禁时间机器），认为时间机器会在产生的瞬间被通过它的强大真空涨落束所毁灭。[14]

伊斯雷尔、泊松和奥里在多罗什克维奇和诺维科夫工作的基础上证明黑洞内的奇点会“变老”；奥里证明当黑洞老而宁静时，落进的物体不会因奇点的潮汐引力而产生严重形变，除非它们到了黑洞的量子引力中心。[13]

夏皮罗和特奥科尔斯基在超级计算机模拟中发现宇宙监督猜想可

能是错误的：裸奇点有可能在非球形星体坍缩中产生。[13]

1993　赫尔塞和泰勒因通过双脉冲星测量证明引力波存在而获诺贝尔奖。[10]

[1]　即《自然哲学之数学原理》，商务印书馆曾出版郑太朴译本；1992年武汉出版社出版了新译本（王克迪译，袁江洋校）。——译者

# 名词

定义和术语

absolute　绝对　独立于任何参照系；每个参照系都观测到相同结果。

absolute horizon　绝对视界　黑洞表面，见视界。

absolute space　绝对空间　牛顿关于我们生活的三维空间的概念，例如绝对静止和绝对长度，即物体长度与测量它的参照系的运动状态无关。

absolute time　绝对时间　牛顿的时间概念，认为时间是普适的，即事件的同时性是惟一的、普遍一致的；两个事件的时间间隔是惟一的、普遍一致的。

accelerated observer　加速观测者　非自由下落的观测者。

accretion disk　吸积盘　包围黑洞或中子星的气体盘。盘内的摩擦力使气体逐渐螺旋下落，被吸积到黑洞或星体。

adiabatic index　绝热指数　同压缩阻抗。

aether　以太　一种假想介质。照19世纪的思想，它在电磁波经过时发生振荡，正是由于它的振荡，波动才表现出来。一般认为以太在绝对空间中静止。

angular momentum　角动量　物体的旋转的度量。在本书里常以旋转来代替“角动量”。

antimatter　反物质　一种物质形式，寻常物质的“对头”。寻常物质的每一类粒子（如电子、质子、中子）几乎总是对应着一种反物质的反粒子（正电子、反质子、反中子）。物质粒子与对应反物质的反粒子相遇时，会彼此湮灭。

apparent horizon　显视界　光子逃出黑洞所遇到的最后引力屏障。当黑洞宁静不变时，显视界与（绝对）视界相同。

astronomer　天文学家　特指用望远镜观测宇宙天体的科学家。

astrophysicist　天体物理学家　特指以物理定律认识宇宙天体行为的物理学家（通常是理论物理学家）。

astrophysics　天体物理学　关于宇宙天体和它们的物理学定律的物理学分支。

atom　原子　物质的基本构成要素。每个原子由带正电的核和周围带负电的电子云构成。电力将电子云束缚在核周围。

atomic bomb　原子弹　通过铀-235或钚-239核裂变链式反应提供爆炸能量的炸弹。

band　频带　频率范围。

bandwidth　带宽仪　器所能探测的频率范围。

bar detector　棒探测器　一种引力波探测器，以传感器监测因波对巨大金属棒的挤压和拉伸而产生的振荡。

beam splitter　分光镜　一种仪器，用以将一束光分裂为沿不同

方向运动的两部分，或者将来自不同方向的光合成为一束。

big bang　大爆炸　宇宙开始的爆炸。

big　crunch　大爆缩或大挤压　宇宙坍缩的终极阶段（假定宇宙最终真会再发生坍缩；我们不知道是否会这样）。

binary system　双星系　两个相互围绕作轨道运动的天体，天体可以是恒星或者黑洞，也可以是一颗恒星和一个黑洞。

BKL　singularity　BKL奇点　周围潮汐引力在时间和空间上随机振荡的奇点。这类奇点可能在黑洞中心或宇宙大爆缩中形成。

black　hole　黑洞　（恒星坍缩形成的）天体，事物可以落进去，但不能逃出来。

black-hole binary　双黑洞系　两个黑洞构成的双星系。

Blandford-Znajek　process　布兰福德-茨纳耶克过程　穿过黑洞的磁场从旋转黑洞汲取旋转能量的过程。

boosted　atomic　bomb　增强原子弹　通过一层或多层聚变燃料增大爆炸威力的原子弹。

chain reaction　链式反应　原子核的系列裂变反应：一次裂变产生的中子触发新的裂变，新裂变中子又触发下一次裂变，等等。

Chandrasekhar　limit　钱德拉塞卡极限　白矮星所能具有的最大质量。

chronology　protection　conjecture　良序猜想或时序保护猜想　霍金关于物理学定律不允许时间机器破坏时间次序的猜想。

classical　经典的　非量子力学的；服从宏观物体的物理学定律的。

cold,dead　matter　冷死物质　所有核反应都完成了的冷物质，耗尽了物质中所有可以利用的核能。

collapsed star　坍缩星　60年代西方对黑洞的称呼。

conservation　law　守恒定律　关于某个特征量永远不会改变的物理学定律。例如质量和能量（通过爱因斯坦的$E=mc^2$统一为一个量）守恒，总电荷守恒和角动量（旋转的总量）守恒。

corpuscle　微粒　17和18世纪对光粒子的称呼。

cosmic censorship conjecture　宇宙监督　猜想关于物理定律不允许物体坍缩时形成裸奇点的猜想。

cosmic　ray　宇宙线　从空间打在地球上的物质或反物质的粒子。有些宇宙线是太阳产生的，但多数也许是从银河系的遥远区域，在超新星喷射到星际空间的热气体云中产生的。

cosmic　string　宇宙弦　一种假想的由空间卷曲产生的一维弦状物体。弦没有端点（它要么像橡皮圈那样自我闭合，要么无限延伸），其空间卷曲使周围的圆的周长除以直径略小于π。

critical　circumference　临界周长　黑洞视界的周长。在此周长以内的物体一定会收缩而形成包围自己的黑洞。临界周长的值是以太阳质量为单位的黑洞或物体质量乘以18.5公里。

curvature of space or spacetime　空间和时空的曲率　空间和时空的一种性质，它偏离欧几里得和闵可夫斯基几何，就是说，它能使平行的直线相交。

Cyg A，即Cygnus A 天鹅A 射电星系，看起来像两个在碰撞的星系（实际上是一个）。这是第一个得到认证的射电星系。

Cyg X-1，即Cygnus X-1 天鹅X-1 银河系中的一个大质量天体，可能是一个黑洞。从地球看，落向它的热气体发射出X射线。

dark star 暗星 18世纪末和19世纪初用来描述我们现在所说黑洞的名词。

degeneracy pressure 简并压力 高密度物质内部由波粒二象性导致的电子或中子的无规则高速运动所产生的压力，物质冷却到绝对零度时，这种压力仍然很强。

deuterium nuclei或deuterons 氘核 一个质子和一个中子在核力束缚下构成的原子核；因氘原子的化学性质与氢相同，所以又叫“重氢”。

differential equation 微分方程 关于函数及其变化率，或者说，关于函数及其“导数”的方程。“解微分方程”的意思是，“根据微分方程计算函数本身”。

Doppler shift 多普勒频移 波源向着接收者运动时，波向高频（短波高能）移动；波源背离接收者运动时，波向低频（长波低能）移动。

electric charge 电荷 物质或粒子产生并感受电力的一种属性。

electric field 电场 电荷周围的力场，吸引或排斥其他电荷。

electric field lines 电力线或电场线 指示电场作用于电荷的力的方向。电力（场）线类似于磁力线。

electromagnetic waves　电磁波　电力和磁力的波，包括波长不同的无线电波、微波、红外辐射、可见光、紫外辐射、X射线和γ射线。

electron　电子　物质基本粒子，带负电荷，遍布原子外层区域。

electron degeneracy　电子简并　电子在高密度下因量子力学波粒二象性而表现的高速无规则运动行为。

elementary particle　基本粒子　物质或反物质的亚原子粒子。其中包括电子、质子、中子、正电子、反质子和反中子。

embedding diagram　嵌入图　为形象表现二维曲面曲率而将它嵌入平直三维空间的一种图。

entropy　熵　大量原子、分子或其他粒子集合的随机性的度量，等于在不改变其宏观表现的条件下粒子分布方式总数的对数。

equation of state　物态方程　物质压力（或压缩阻抗）对密度的依赖关系。

equivalence principle或principle of equivalence　等效原理　在有引力的局部惯系参照系中，物理学定律应与在没有引力的惯性参照系中具有相同形式。

error box　误差区间　观测确定的某一特定恒星或天体在空间所处的范围。观测的不确定性（误差）越大，这个区间就越大。

escape velocity　逃逸速度　为了摆脱物体的引力作用，从这个引力体表面发射的物体所必须具有的速度。

event　事件　时空中的一点，即特定时刻的一个空间位置。也可以说是发生在时空某一点的事物，如鞭炮爆炸。

exotic　material　奇异物　以近光速经过它的某个观测者所测得的能量密度是负的。

field　场　连续光滑分布于空间的事物。如电场、磁场、时空曲率和引力波。

fission,nuclear　核裂变　大原子核分裂为几个较小的核。铀核或钚核的裂变是原子弹爆炸的能源，裂变也是核反应堆的能源。

freely　falling　object　自由落体　除引力外不受其他力作用的物体。

free particle　自由粒子　不受力作用的粒子，即只在自身惯性影响下运动的粒子。引力出现时，指除引力外不受其他力作用的粒子。

frequency　频率　波的振荡速率，即每秒钟的振荡次数。

frozen star　冻星　60年代苏联对黑洞的称呼。

function　函数　说明一个量如何依赖于其他量的数学表述形式，如黑洞视界的周长与黑洞质量的依赖关系是$C=4\pi CM/c^2$，这里C为周长，M为质量，G是牛顿引力常数，c是光速。

nuclear fusion　核聚变　两个小原子核形成一个较大的核。太阳的热量和氢弹的动力都来自氢、氘和氚核聚变形成氦核。

galaxy　星系　围绕某个共同中心的10亿到1万亿颗恒星的集合。星系的典型大小是100 000光年。

gamma rays　γ射线　极短波长的电磁波。参见图P.2。

Geiger counter　盖革计数器　德国物理学家盖革（Geiger,Hans Wilhelm）1913年发明的一种探测X射线的简单仪器，也叫“正比计数器”。

general relativity　广义相对论　爱因斯坦将引力描写为时空曲率的物理学定律。

geodesic　测地线　弯曲空间或弯曲时空中的直线，也叫短程线。在地球表面，测地线就是大圆。

gigantic black hole　巨黑洞　比1百万或更多太阳质量还重的黑洞，如我们认为处在星系和类星体中心的黑洞。

global methods　整体方法　以拓扑学与几何学的结合为基础的分析时空结构的数学方法。

gravitational cutoff　引力隔绝　奥本海默用来描述包围坍缩恒星的黑洞的形成的名词。

gravitational lens　引力透镜　黑洞、星系等引力体通过偏转来自遥远光源的光线而使其聚焦的作用。参见光线偏折。

gravitational redshift of light　光的引力　红移光线离开引力场时波长被拉长（颜色变红）。

gravitational time dilation　引力时间膨胀　引力源附近时间流变慢。

gravitational wave　引力波　以光速传播的时空曲率波。

graviton　引力子　根据波粒二象性与引力波相联系的粒子。

gyroscope　陀螺　能长时间稳定旋转轴向的快速旋转体。

“hair” 毛 黑洞将通过辐射失去的一切性质，例如磁场和视界面上的隆起。

hoop conjecture 环猜想 索恩提出的猜想：当且仅当物体被压缩到能在任何方向套进临界周长的环时，它才能形成黑洞。

horizon 视界 黑洞表面，没有东西能离开它返回。也叫绝对视界，以区别于显视界。

hydrogen bomb 氢弹 爆炸能量来自氢、氘、氚核形成氦核的聚变的核弹。参见超弹。

hyperspace 超空间 一种假想的平直空间，我们想象宇宙的弯曲空间碎片就嵌在其中。

implosion 坍缩星体 在自身引力作用下的高速收缩。

inertia 惯性 物体对作用于它的力的加速的抵制作用。

inertial reference frame 惯性参照系 既不转动也无外力作用的参照系，它只在自身惯性作用下运动。参见局部惯性参照系。

infrared radiation 红外辐射 波长比可见光略长的电磁波。见图P.2。

interference 干涉 两波的线性叠加表现。两波峰谷同步时，干涉使波增强（相长干涉或结构干涉）；峰谷相反时，干涉使波减弱（相消干涉或破坏性干涉）。

interferometer 干涉仪 以波的干涉为基础的仪器。见射电干涉仪和干涉仪探测器。

interferometric detector 干涉仪探测器 一种引力波探测器。

利用激光束的干涉来监测引力波潮汐力产生的悬挂物体的运动。也叫干涉仪。

interferometry　干涉测量法　多个波的干涉过程。

intergalactic space　星系际空间　星系间的空间。

interstellar space　星际空间　银河系里恒星间的空间。

inverse square law of gravity　引力的平方反比定律　牛顿的引力定律。宇宙中任何一对物体间都存在使它们彼此靠近的吸引力，它正比于物体质量的积而反比于物体间距离的平方。

ion　离子　失去部分轨道电子从而带正电荷的原子。

ionized gas　电离气体　大部分组成原子失去轨道电子的气体。

jet　喷流　从射电星系或类星体中心喷向遥远射电叶的高能气流。

laws of physics　物理学定律　通过逻辑和数学运算能导出宇宙行为的基本原理。

length contraction　长度收缩　经过观测者的物体在运动方向上发生的长度收缩。

light　光　人眼能见的电磁波。见图P.2。

light deflection　光偏转　光或其他电磁波经过太阳或其他引力体时，由于物体周围的时空曲率而发生的传播方向的偏转。

LIGO　引力波的激光干涉仪观测（The Laser Interferometer Gravitational-Wave Observatory）。

linear　线性　通过简单加法而合成的性质。

lobe　叶星　系或类星体外的巨大射电气体云。

local inertial reference frame　局部惯性参照系　除引力外不受其他力作用的参照系，在引力作用下自由下落，而且小到可以忽略内部的潮汐引力的速度。

magnetic field　磁场　产生磁力的场。

magnetic field lines　磁力线　现在称磁感应线。指示磁场方向（即磁场中的罗盘指针的方向）的线。在磁棒上面放一张纸，在纸上洒些铁粉，就能显现磁棒的力线。

mass　质量　物体的物质的量度。（物体惯性正比于质量，爱因斯坦证明质量实际上是能量的紧致形式。）在物体的惯性很重要时，“质量”也用来指“质量形成的物体”。

Maxwell's laws of electromagnetism　麦克斯韦电磁学定律　J.C.麦克斯韦统一所有电磁现象的一组定律。根据这些定律，可以用数学方法导出电、磁和电磁波的行为。

metaprinciple　形而上原理　一切物理学定律都应服从的原理。例如，相对性原理。microsecond　微秒　百万分之一秒。

microwaves　微波　波长比无线电波略短的电磁波；见图P.2。

Milky Way　银河　我们所在的星系。

mixmaster singularity　搅拌式奇点　在这种奇点附近，潮汐引力随时间混沌地振荡，但在空间上不一定变化。参见BKL奇点。

molecule　分子　几个共享电子云的原子结合而成的实体。例

如，水分子就是由两个氢原子和一个氧原子构成的。

mouth　洞口　虫洞的出入口，在虫洞的两端。

naked　singularity　裸奇点　不在黑洞内的奇点（即不被黑洞视界包围），因而外面的人可以看到并研究它。参见宇宙监督猜想。

National　Science　Foundation（NSF）　国家科学基金会　美国政府负责资助基础科学研究的办事机构。

natural　philosopher　自然哲学家　17、18和19世纪广泛使用的对我们现在所说的科学家的称呼。

nebula　星云　星际空间中明亮的发光气体云。30年代以前，星系曾被普遍误认为星云。

neutrino　中微子　一种类似于光子的极轻粒子，但几乎不与物质发生作用。例如，太阳中心产生的中微子几乎不会被吸收和散射而从太阳周围的物质中飞出来。

neutron　中子　一种亚原子。中子和质子由核力束缚在一起形成原子核。

neutron　core　中子核　奥本海默为中子星取的名字。也指正常恒星中心的中子星。

neutron　star　中子星　一种约一个太阳质量但周长只有50到1000公里的星体，由紧密堆积的中子在引力作用下形成。

new quantum mechanics　新量子力学　1926年建立的量子力学定律的最终形式。

Newtonian　laws　of　physics　牛顿物理学定律　在牛顿绝对空间

和时间概念基础上建立的物理学定律，是19世纪宇宙思想的核心。

Newton's law of gravity　牛顿的引力定律　见引力的平方反比定律。

no-hair conjecture　无毛猜想　60和70年代提出（70和80年代证明）的一个猜想：黑洞的一切性质由它的质量、电荷和旋转惟一决定。

nonlinear　非线性　不能通过简单加法而由更复杂的方式相结合的性质。

nova　新星　老恒星突然的光爆发，现在知道它由恒星外层核爆炸引起。

nuclear burning　核燃烧　恒星热量和氢弹的能量来源的核聚合反应。

nuclear force　核力　也叫“强相互作用”。质子与质子、质子与中子和中子与中子之间的力，将它们约束在一起形成原子核。粒子相互远离时，核力是吸引力；粒子靠近时，它是排斥力。核力对中子星中心附近的压力有重大影响。

nuclear reaction　核反应　几个原子核结合成一个更大的核（聚变），或较大的一个核分裂为几个更小的核（裂变）。

nuclear reactor　核反应堆　用以进行核裂变链式反应的设施，产生能量或钍，有时也用来发电。

nucleon　核子　中子或质子。

atomic nucleus　原子核　原子的致密核心。原子核由中子和质

子在核力约束下构成，带正电。

observer　观察者　进行测量的（通常是假想的）人。

old quantum mechanics　旧量子力学　20世纪头一二十年发展起来的量子力学早期形式。

optical astronomer　光学天文学家　利用可见光（人眼能见的光）观测宇宙的天文学家。

orbital period　轨道周期　物体沿轨道绕另一物体完成一周所用的时间。

paradigm　规范或范式　科学家群体研究某一问题并在彼此间交流研究结果所采用的一系列工具和方法。

particle　粒子　小实体，构成物质的要素（如电子、质子、光子和引力子）。

perihelion　近日点　行星轨道距太阳最近的那一点。

perihelion shift of mercury　水星近日点移动　水星椭圆轨道不能完全自我封闭，结果表现为水星每次经过近日点时，它都略有移动。

perturbation　扰动　物体或它周围的弯曲时空（离开正常形态）的微小扭曲变形。

perturbation methods　微扰法　用以分析物体（如黑洞）小扰动行为的数学方法。

photon　光子　光或其他类型电磁波（无线电波、微波、红外、紫外、X射线、γ射线）的粒子；根据波粒二象性，这种粒子即是与电

磁波相关联的粒子。

piezoelectric crystal 压电性晶体 在压缩或拉伸时产生电压的晶体。

Planck's constant 普朗克常数 记为$\hbar$，量子力学定律基本常数；光子的能量与其角频率（即2π乘以频率）之比，$1.055\times 10^{-27}$尔格·秒。

Planck-Wheeler length,area,time 普朗克-惠勒长度、面积、时间 与量子引力相关的量。普朗克-惠勒长度，$\sqrt{G\hbar/c^3}=1.62\times 10^{-33}$厘米，据我们所知，在这一长度尺度下，空间不再存在，而成为量子泡沫。普朗克-惠勒时间（1/c乘以普朗克-惠勒长度，约$10^{-43}$秒）是可能存在的最短时间间隔；如果两个事件间隔更小，则不能分辨其先后次序。普朗克-惠勒面积（普朗克-惠勒长度的平方，即$2.61\times 10^{-66}$平方厘米）在黑洞的熵中起着关键作用。在以上公式中，$G=6.670\times 10^{-8}$达因·厘米$^2$/克$^2$是牛顿引力常数，$\hbar=1.055\times 10^{-27}$尔格·秒是普朗克量子力学常数，$c=2.998\times 10^{10}$厘米/秒是光速。

plasma 等离子体 电离的热导电气体。

plutonium-239 钚-239 一类特殊的钚原子核，包括239个核子（94个质子，145个中子）。

polarization 极化 电磁波和引力波由在不同方向振荡的两个分量构成，这两个分量即被称为波的两个极化。

polarized body 极化体 正负电荷分别聚集在不同区域的物体。

polarized light；polarized gravitational waves 极化光；极化引力波 一种极化完全不出现（消失了）的光波或引力波。

postdoc　博士后　博士后研究者。刚获博士学位者继续在更高级的研究者指导下从事科研训练。

pressure　压力　物质受压时产生的向外的力的总和。

Price's　theorem　普赖斯定理　所有能转化为辐射的黑洞性质都将转化为辐射，而且终将被完全辐射，从而黑洞成为“无毛”的。

primordial　black　hole　原生黑洞　在大爆炸中产生的质量通常远远小于太阳的黑洞。

principle of absoluteness of the speed of light　光速的绝对性原理　爱因斯坦在狭义相对论中提出的一个原理：光速为一普适常数，在所有方向和每个惯性系中都是一样的，与参照系的运动无关。

principle　of　equivalence　等效原理　在引力作用下的局部惯性参照系中的所有物理学定律都应该与它们在没有引力的惯性参照系中一样。

principle　of　relativity　相对性原理　爱因斯坦在狭义相对论中提出的一个原理：物理学定律不能区分不同的惯性参照系；就是说，它们在每个惯系参照系中应该有一样的形式。引力出现时，局部惯性参照系起着原理中惯性参照系的作用。

pulsar　脉冲星　发出辐射束（射电波，有时也有光和X射线）的磁化的旋转中子星。星体旋转时，辐射束像旋转的聚光灯束那样扫过地球，每扫过一次，天文学家就收到一个辐射脉冲。

pulsation　脉动　物体（如黑洞、星体或铃铛）的振动或振荡。

quantum　field　量子场　量子力学定律所决定的场。以足够的精度测量时，所有场都表现为量子场；但在一般测量精度下，它们可能

表现出经典行为（即不表现波粒二象性和真空涨落）。

the laws of quantum fields in curved spacetime　弯曲时空的量子场定律　广义相对论（弯曲时空）与量子场定律的部分结合，其中引力波和非引力场作为量子力学的，而它们所在的弯曲时空还是经典的。

quantum　foam　量子泡沫　一种概率的泡沫或空间结构，可能构成奇点中心，也可能在普朗克-惠勒长度或更小尺度下出现在普通空间。

quantum　gravity　量子引力　结合广义相对论与量子力学而得到的物理学定律。

quantum　mechanics　量子力学　主宰微观领域（分子、原子、电子、质子）的物理学定律，也是宏观领域的基础，但很少表现出来。量子力学预言的现象，有测不准原理、波粒二象性和真空涨落等。

quantum　nondemolition　量子无破坏　克服标准量子极限的一种测量方法。

quantum theory　量子理论　同量子力学。

quasar　类星体　遥远宇宙中的致密高光亮天体，可能是巨黑洞的能源。

radiation　辐射　任何形式的高速粒子或波。

radio　astronomer　射电天文学家　根据射电波研究宇宙的天文学家。

radio galaxy　射电星系　发射强大电波的星系。

radio interferometer　射电干涉仪　由几个射电望远镜连结而构成的实验设施，作用像一个特大的射电望远镜。

radio source　射电源　发射电波的任意天体。

radio telescope　射电望远镜　通过电波观测宇宙的望远镜。

radio waves　无线电波　低频电磁波，人类用来传播广播信号，天文学家用以研究遥远天体。见图P.2。

redshift　红移　电磁波向波长更长，即颜色“更红”的波转移。

reference frame　参照系　一个（可能是假想的）在宇宙中以某种方式运动的进行物理学测量的实验室。

relative　相对　对参照系的依赖；在宇宙中不同运动方式的参照系的测量不同。

resistance to compression　压缩阻抗或阻抗（resistence）　也叫绝热指数。物质内部密度增加1个百分点所对应的压力增加的百分点。

rigor；rigorous　严格　高度的精确性和可靠性（用于数学计算和论证）。

rotational energy　旋转能　与黑洞或恒星或其他物体的旋转相关联的能量。

Schwarzschild geometry　史瓦西几何　非旋转球状黑洞内部和周围的时空几何。

Schwarzschild singularity　史瓦西奇点　在1916到约1958年间用来称我们今天所说的黑洞。

ScoX-1，即Scorpius X-1　天蝎X-1　天空中最亮的X射线星。

second law of thermodynamics　热力学第二定律　熵永不减少而几乎总是增加。

sensitivity　灵敏度　某仪器所能测量的最弱信号，即仪器测量信号的能力。

sensor　传感器　监测棒的振动或物质运动的仪器。

shocked gas　受激气体　在激波前沿被加热和压缩的气体。

shock front　激波前沿　气流中气体密度和温度突然剧烈跃升的地方。

simultaneity breakdown　同时性的丧失　在一个参照系中测量为同时的事件在另一个参照系看来是不同时的。

singularity　奇点　广义相对论破灭而量子引力发生作用的曲率极大的时空区域。如果只用广义相对论来描写奇点，我们会（错误地）看到，那里的引力和时空曲率是无限大的。量子引力可能会以量子泡沫来取代这些无限。

Sirius B　天狼B　天狼星的白矮星伴星。

spacetime　时空　空间和时间的四维统一“结构”。

spacetime curvature　时空曲率　原先沿平行世界线运动的自由下落粒子将因这种时空特性而靠拢或分离。时空曲率与潮汐引力是同一事物的不同名称。

spacetime diagram　时空图　以时间为纵坐标、空间为横坐标的图。

special relativity　狭义相对论　爱因斯坦在无引力条件下创立的一组物理学定律。

spectral　lines　光谱线　某光源所发的光在光谱上表现的鲜明特征。这些特征源于特定原子或分子强烈发射特定波长的波。

spectrograph　摄谱仪　一种复杂的棱镜，用以分解不同颜色（波长）的光从而测量其光谱。

spectrum　光谱　电磁波存在的波长或频率范围，从极低频的无线电波到光到极高频的$\gamma$射线。见图P.2。也可以表现为光作为频率（或波长）的函数的分布图像，这可以让光通过棱镜而得到。

spin或rotation　旋转　见角动量。

stability　稳定性　关于物体是否稳定的问题。参见不稳定。

standard　quantum　limit　标准量子极限　测不准原理产生的标准测量方法的确定性极限；可以通过量子无破坏方法来克服。

stroboscopic　measurement　频闪测量　一种特殊的量子无破坏测量方法，即对振荡棒做一系列快速测量，每次测量间隔一个振动周期。

structure of a star　星体结构　星体压力、密度、温度和引力随到中心距离的变化情况。

superbomb　超弹　能产生任意大爆炸的氢弹。

superconductor　超导　无任何电阻的理想导电材料。

supermassive star　超大质量恒星　比10 000个太阳还重的假想恒星。

supernova　超新星　死星的巨大爆发。星体外层爆发的能量来自内核坍缩成中子星时的能量释放。

surface gravity　表面引力　粗略地讲，即正好静止在黑洞视界上方的观测者所感受的引力作用的强度。（更准确说，是那个引力乘以观测者所在位置的引力时间膨胀。）

synchrotron radiation　同步辐射　绕磁力线螺旋运动的高速电子发射的电磁波。

thermal pressure　热压力　原子、分子或其他粒子的随机热运动产生的压力。

thermodynamics　热力学　关于大量原子、分子及其热量的随机统计行为的一套物理学定律。

thermonuclear reactions　热核反应　热导致的核反应。

tidal gravity　潮汐引力　在不同方向拉伸或压缩物体的引力加速度。月亮和太阳的潮汐引力在地球上引起海洋潮汐。

time dilation　时间膨胀　时间流（因引力作用）变慢的效应。

time machine　时间机器　回到过去的旅行设计，用物理学术语，即“闭合类时曲线”。

topology　拓扑　研究物体相互连结或自我连结的定性方式的数学分支。例如，球（无洞）与圈（有一个洞）具有不同的拓扑。

tritium　氚　即超重氢，由1个质子和2个中子在核力束缚下形成

的原子核。

ultraviolet radiation　紫外辐射　波长比光略短的电磁辐射，见图P.2。

uncertainty principle　测不准原理　量子力学基本定律之一。如果以很高的精度测量物体的位置或场的强度，则这样的测量必然会对物体的速度或场的变化率产生不可预料的干扰。

universe　宇宙　与空间所有其他区域分离的一个空间区域，就像一个岛屿同别的陆地分离一样。也指我们所在的宇宙。

unstable　不稳定　物体受轻微扰动时，扰动会增大，从而巨大地改变物体状态，甚至破坏它。用更复杂的术语说，即“对微小扰动是不稳定的”。

uranium-235　铀-235　铀核的一种，含92个质子和143个中子。

vacuum　真空　所有粒子、场和能量都尽可能消除了的时空区域，只留下不可消除的真空涨落。

vacuum fluctuations　真空涨落　通过空间小区域间的瞬时能量“交流”而产生的随机的、不可预测、不可消除的场（如电磁场或引力场）振荡。参见真空和虚粒子。

virtual particles　虚粒子　利用从相邻空间获得的能量而成对产生的粒子。量子力学要求其能量必须立刻还原，于是虚粒子很快湮灭而不能被捕获。在自由下落的观测者看来，虚粒子是真空涨落的粒子表现。虚光子和虚引力子分别是电磁和引力的真空涨落的粒子表现。参见波粒二象性。

warpage of spacedme　时空弯曲　同时空曲率。

wave　波　场（如电磁场或时空曲率）在时空中传播的振荡。

waveform　波形　说明波的振荡细节的曲线。

wavelength　波长　波的两峰（或谷）间的距离。

wave-particle　duality　波粒二象性　一切波有时表现为粒子，一切粒子有时表现为波。

white-dwarf star　白矮星　耗尽了所有核燃料并逐渐冷却的大约为太阳质量和地球周长的星体。它通过电子简并压力抵抗自己引力的挤压。

world line　世界线　物体在时空或时空图上的路径。

wormhole　虫洞　连结在我们宇宙中相隔遥远的两个位置的拓扑空间里的一个“柄”。

X-rays　X射线　波长在紫外辐射和γ射线之间的电磁波。见图P.2。

# 注释

我凭什么相信

我说的？

材料来源和缩写

注释中引用的材料列在参考文献中。

注释中使用的缩写记号是

ECP-1—The Collected Paper of Albert Einstein,Volume,1（爱因斯坦全集，第1卷），参考文献中引作ECP-1。

ECP-2—The Collected Paper of Albert Einstein,Volune,2（爱因斯坦全集，第2卷），参考文献中引作ECP-2。

INT—作者所做采访，列在参考文献开头。

MTW—Misner,Thorne and Wheeler（1973）。[1]序幕1.这段话改自Thorne（1974）。

2.牛顿公式为$M_h=C_0^3/(2\pi P_0^2)$，其中$M_h$为黑洞（或任何其他引力体）的质量，$C_0$和$P_0$为围绕黑洞的圆形轨道周长和周期，$\pi=3.14159\ldots$，G为牛顿引力常数，$1.327\times10^{11}$千米$^3$/秒$^2$·太阳质量。参见下面第1章注4。将飞船轨道周期$P_0=5$分46秒和周长$C_0=10^6$千米代入公式，得$M_h=10$太阳质量。（1个太阳质量

$M_\odot = 1.989 \times 10^{30}$千克。）（相同符号代表相同意思的，以下不再说明——译者）

3.视界周长公式为$C_h = 4\pi GM_h/c^2 = 18.5 \times$（$M_h/M_\odot$）千米，其中$c = 2.998 \times 10^5$千米/秒为光速。参见，如MTW第31、32章。

4.以头脚间（或任意两个物体间）相对加速度表示潮汐力为$\triangle\alpha = 16\pi^3 G$（$M_h/C^3$）L，这里C为你所在位置的轨道周长，L为头脚间距离。注意1个地球引力是$g = 9.81$米/秒$^2$。例如，见MTW，p.29。

5.上面（注4）的公式给出潮汐力$\triangle\alpha \propto M_h/C^3$，周长接近视界时，$C \propto M_h$，于是$\triangle\alpha 1/M_h^2$。

6.飞船时间$T_s$、地球时间$T_e$和旅行距离D的关系是$T_e =$（2c/G）sinh（$gT_s/2c$）和$D =$（$2c^2/g$）〔cosh（$gT_s/2c$）-1〕，这里g是飞船的加速度（等于“1个地球引力”，9.81米/秒$^2$），cosh和sinh分别为双曲余弦和双曲正弦函数，例如见MTW第6章。如果旅行时间远大于1年，则公式近似为$T_e = D/c$和$T_s =$（2c/g）In（$gD/c^2$），这里In为自然对数。

7.对围绕非旋转黑洞圆形（或其他）轨道的数学分析可以看MTW第25章，特别是其中的卡片25.6。

8.在质量为$M_h$、视界周长为$C_h$的黑洞上方周长为C的轨道上，你感觉的加速度为$\alpha = 4\pi^2 G$（$M_h/C^2$）/ $\sqrt{1 - C/C_h}$。如果接近视界，那么$C \approx C_h \propto M_h$，意味着$\alpha \propto 1/M_h$。

9.见上面注6。

10.当观测者在周长为$C_h$的视界上方周长为C的轨道上时，他看到所有从外面宇宙来的光都聚在一个角直径$\alpha \approx 3\sqrt{3}\sqrt{1 - C/C_h}$弧度

≈300 $\sqrt{1-C/C_h}$度的明亮圆盘里。例如，参见NTW卡片25.7。

11.当观测者在视界上方周长为C的轨道上时，他看到所有从外面宇宙来的光的波长λ都将蓝移（引力红移的反方向移动），大小是$\lambda_{接收}/\lambda_{发射}=1/\sqrt{1-C/C_h}$。例如，见MTW p.657。

12.两个质量为$M_h$的相距为D的相互环绕的黑洞的轨道周期为$2\pi\sqrt{D^3/2GM_h}$，引力波反冲使它们在（5/512）×（$c^5/G^3$）×（$D^4/M_h^3$）的时间后螺旋式靠近并结合。见MTW方程（36.17b）。

13.大梁环上距中心层距离L的人感到向中心的加速度$\alpha=(32\pi^3GM_h/C^3)L$，这是由旋转环的离心力和黑洞潮汐力作用的结果。这里C是环中心的周长。可以拿它与地球的引力加速度9.81米/秒$^2$比较。

14.$10^{-33}$厘米＝$\sqrt{G\hbar/c^3}$为“普朗克-惠勒长度”。$\hbar$为普朗克常数（$1.055\times10^{-34}$千克·米$^2$/秒）。参见第14章。

15.例如，可以看Will（1986）。

第1章

一般说明：本章中多数关于爱因斯坦生平的材料来自几部关于他的标准传记：Pais（1982）[2]，Hoffman（1972），Clark（1971），Einstein（1949）和Frank（1947）。从这几本书里引用的多数历史评述在下面都不再单列了。现在，正在陆续出版的爱因斯坦文集ECP-1，ECP-2以及Einstein and Marić（1992），提供了许多新的历史资料，下面引用了不少它们的东西。

1　ECP-1,Document 99。

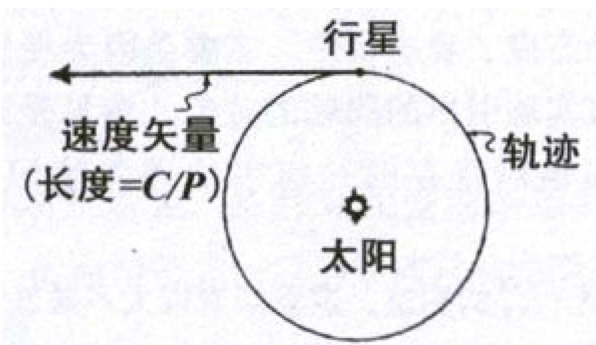
2　ECP-1,Document 115,Renn and Schulmann（1992）第xix页

有[英]译文。

3　下面的例子说明了什么是物理学定律的“数学操作”：

17世纪初，开普勒（Johannes　　　Kepler）根据第谷（Tycho Brahe）的行星观测数据导出行星轨道周长C的立方除以轨道周期P的平方，即$C^3/P^2$，对所有已知行星（水星、金星、地球、火星、木星和土星）都是常数。半个世纪后，牛顿通过他的运动和引力定律的“数学操作”，解释了开普勒的发现：

（1）根据下面的图，动动脑筋，可以导出行星环绕太阳时速度的变化率为$2\pi C/P^2$，它有时被称为轨道行星的离心加速度。




（2）牛顿第二运动定律指出，这样的速度变化率（离心加速度）必须等于太阳对行星的引力作用$F_g$除以行星质量$M_P$，也就是$2\pi C/P^2=F_g/M_P$。

（3）牛顿引力定律告诉我们，引力$F_g$正比于太阳质量$M_s$乘以行星质量$M_P$除以行星轨道周长的平方，写成等式即$F_g=4\pi^2 GM_sM_s/C^2$，这里G为牛顿引力常数，$6.67\times 1O^{-20}$千米$^3$/秒$^2$·千克，或者$1.327\times 1O^{11}$千米$^3$/秒$^2$太阳质量。

（4）将引力表达式代入牛顿第二定律，得$2\pi C/P^2=4\pi^2 GM_s/$

$C^2$，方程两端同乘以$C^2/2\pi$，得$C^3/P^2=2\pi GM_s$。

这样，牛顿的运动定律和引力定律解释了——实际上也可以说它们要求——开普勒所发现的关系：$C^3/P^2$对所有行星都相同；它只与牛顿引力常数和太阳质量有关。

为说明物理学定律的威力，我们看上面的运算不仅解释了开普勒的发现，而且为我们提供了一个称量太阳的方法。在（4）中最后一个等式两端同除以$2\pi G$，我们得到太阳质量的公式$M_s=C^3/(2\pi GP^2)$。将天文学家观测的任意行星轨道的周长C和周期P，以及物理学家在地球实验室里测得的牛顿引力常数G代入这个公式，我们得出太阳质量为$1.989\times10^{30}$千克。

4　ECP-1,Document 39；Einstein and Marić（1992），Document 2。

5　19世纪末叶，某些物理学家猜测地球附近的以太可能会被地球在绝对空间的运动所拖曳，我在本章没有谈这一点。实际上，实验证据有力地否定了这种牵引作用；假如在地球表面附近以太相对于地球静止，那么就不应该有星光的光行差；而地球环绕太阳引起的光行差是确凿的事实。关于以太的思想史的简单讨论可以参考Pais（1982）第6章；更详细的讨论，可以看那里所引的文献。

6　迈克尔逊时代的技术还不足以精确（$1/10^4$的精度）比较不同方向的单程光速以检验牛顿的预言。不过，对回路光速也有类似预言（平行于地球在以太中的运动方向上的光速，与垂直方向的，大约存在$5/10^9$的差异），迈克尔逊的技术对测量这种回路差别是很理想的；迈克尔逊寻找这种差别，但是没能找到。因此不同方向的光速应该是一样的。

7　我不能肯定，也许韦伯相信这一点，但觉得在课堂上讲迈克

尔逊-莫雷实验不太恰当才特别表现出那种态度。我这么想，主要是因为在韦伯的讲课中，我没有发现任何讨论这一实验或实验引出的问题的迹象。参见爱因斯坦的听课笔记（ECP-1，Document 37）和他对仅存的一组韦伯课堂笔记的简单描述（ECP-1，p.62）。

8 另一些实验包括对星光光行差的测量，这意味着以太不被地球所拖曳。见上面注5。

9 想一下（注6），迈克尔逊实际是在测量回路光速，在寻找约10亿分之5的随方向的变化。

10 “磁力线没有端点”的讨论和图1.1中更详细的讨论，是我用现代的形象语言表达的麦克斯韦方程的一个方面；关于洛伦兹、拉莫和彭加勒与这个问题的详细讨论，见Pais（1982），p.123～130。

11 为将这些定律表述得更优美，不仅需要运动物体的长度收缩和时间膨胀，还需要假定同时性是相对的，即同时性依赖于观测者的运动状态；洛伦兹、拉莫和彭加勒不仅对长度收缩和时间膨胀很关心，也同样非常重视这个问题。不过，为了说得简单明白，我在正文里没有讲这些；到第1章后面才考虑了同时性问题。

12 ECP-1,Document 52；Einstein and Marić（1992）,Document 8。

13 我又在这儿猜测了。我们并不真正了解爱因斯坦在1899～1905年间在多大程度上考虑这些问题。Pais（1982,6b）说明，在那6年里，爱因斯坦不知道洛伦兹-彭加勒-拉莫根据麦克斯韦定律所做的长度收缩和时间膨胀的推论。更准确说，他知道洛伦兹的一阶速度的洛伦兹变换的结果（包括同时性的破灭），但不知道出现长度收缩和时间膨胀的二阶结果。另一方面，他可能知道菲兹杰拉德和洛伦兹根据迈克尔逊-莫雷实验做出的长度收缩的推测。我们倒是知道，他在

1905年关于狭义相对论的文章里提出了自己的完全的洛伦兹变换的推导，在所有阶都是准确的；他还自行推导了长度收缩、时间膨胀和同时性的破灭。

14关于玛丽奇个性的描述，主要根据她和爱因斯坦的情书，见Renn and Schulmann（1992）；情书见ECP-1或Einstein and Marić（1992）。

15 ECP-1，Document 94；Einstein and Marić（1992），Document 95。

16 ECP-1，Document 100。

17 ECP-1，Document 138。

18 ECP-1，Document 125。

19 ECP-1，Document 104。

20 ECP-1；Renn and Schulmann（1992）；Einstein and Marić（1992）。

21 根据多种爱因斯坦传记，我猜他的大多数自由时间都是这样度过的。

22 Seelig（1956），转引自Clark（1971）。

23 关于贝索对爱因斯坦工作的影响，还是去看Renn and Schulmann（1992）第xxvi页的讨论。

24 ECP-2,Document 23，Section 2。

25 例如，参见Will（1986）附录。

26　正如Pais（1982，6b.6）所说，彭加勒比爱因斯坦早一年建立了相对论原理（叫“相对性原理”）的原始形式，但不知道它的作用。

27　ECP-2，Document 23。

第2章

一般说明：本章多数关于爱因斯坦生平的材料来自几部他的标准传记：Pais（1982）,Hoffman（1972），Clark（1971），Einstein（1949）和Frank（1947）。从这几本书里引用的多数历史评述在下面都不再单列。在未来的几年，随着爱因斯坦文集的陆续出版（已经出版了ECP-1和ECP-2），还会出现一些新的材料。

爱因斯坦从狭义相对论走向广义相对论的认识路线，基本上就像本章描述的那样。不过，我还是做了些必要的较大的简化；为了清楚说明他的路线，我用了现代的物理学语言，而没有用爱因斯坦当年用过的。关于爱因斯坦思想路线的历史重建，请看Pais（1982）。

1　闵可夫斯基的讲话，是1908年9月21日在科隆举行的第80届德国自然科学家和医生大会上发表的，它的英译本发表在Lorentz,Einstein,Minkowski and Weyl（1923）。[3]

2　月亮在绕地球的运动中好像有很小的加速，这是牛顿引力定律不能解释的现象。1920年，泰勒（G.I.Taylor）和杰弗瑞斯（H.Jeffries）认识到，月亮实际并没有在加速，倒是地球的自转因为月亮的引力对海洋的潮汐作用而变慢了。天文学家将月亮的稳定运动与地球变慢的自转比较，便错误地推测月亮快了。见Smart（1953）。

3　爱因斯坦这篇优美的综述文章的英译本是ECP-1，Document 47。

4　卡片2.4的爱因斯坦论证原发表在Einstein（1911）。

5　ECP-2，Document 47。

6　见Frank（1947），p.89～91。

7　Einstein（1915）。

8　卡片2.6：为熟悉广义相对论数学形式的读者做些说明：卡片里的爱因斯坦场方程对应的数学关系是$R_{tt}=4\pi G(T_{tt}+T_{xx}+T_{yy}+T_{zz})$这里$R_{tt}$是里奇张量的时间-时间分量，G是牛顿引力常数，$T_{tt}$是以能量单位表示的质量密度（见卡片5.2），$T_{xx}+T_{yy}+T_{zz}$为沿三个相互垂直方向的主应力（压力）之和。见MTW p.406，爱因斯坦场方程的“时间-时间”分量在一切参照系都成立时，也保证了其他9个分量成立。

9　爱因斯坦的个人文稿和部分发表文章的版权已聚讼几十年了。苏联出版他的俄文版文集时还没有签署国际版权协定。更完备的爱因斯坦文集正在陆续出版，头两卷即ECP-1和ECP-2。[4]

第3章

1　Einstein（1939）。

2　Michell（1784）。这一工作的讨论见Gibbons（1979），Schaffer（1979），Israel（1987）和Eisenstaedt（1991）。

3　Laplace（1796,1799）。Isrel（1987）和Eisenstaedt（1991）讨论了拉普拉斯关于暗星的出版物。Eisenstaedt讨论了那时通过观测证明米歇尔预言（即大质量恒星发出的光受恒星引力作用）所进行的努力和失败，还认为可能就是因为这些失败，拉普拉斯才在他著作的第3版中删除了暗星的讨论。

4　Schwarzschild（1916a,b）。

5　Brault（1962）。关于爱因斯坦广义相对论引力定律的检验的详细讨论，请看Will（1986）。

6　关于人们对史瓦西几何的反应和研究的早期历史，Eisenstaedt（1982）有详细讨论；从1916到1974年的更粗线条的历史，可以在Israel（1987）中找到。

7　Einstein（1939）。

8　Israel（1916b）。

9　Israel（1990）。

10　Israel（1990）。

第4章

一般说明：本章的历史记述主要根据（i）过去25年来与钱德拉塞卡的个人谈话，（ii）对他的录音访问（INT-Chandrasekhar），（iii）他写的一本关于爱丁顿的书（Chandrasekhar,1983a），和（iv）关于他的一本优秀传记（Wali，1991）。除特殊情况，我不再对具体的材料说明来源。钱德拉塞卡关于白矮星的科学著作收在Chandrasekhar（1989）。

1　Fowler（1926）。

2　Eddington（1926）。

3　脚注：关于亚当斯遇到的困难和他在测量中犯的错误，请看Greenstein，Oke，和Shipman（1985）的详细讨论。这篇文献还提供了1985年以前对天狼B的观测研究情况。

4　这里，我大胆写了两种情况。第一，Fowler（1926）已经计算了压缩阻抗，所以钱德拉塞卡只需检验他的计算；第二，钱德拉塞卡的计算并没沿这条路线（INT-Chandrasekhar），虽然它与他实际用的方法在数学上是等价的。对我来讲，那条路线是最容易解释的，而实际的计算需要在电子动量空间上进行压力积分。

5　Chandrasekhar（1931）。

6　脚注：Stoner（1930）。Chandrasekhar（1931）简单提到了Stoner的工作。Stoner的研究和Wilhelm　Anderson的相关研究的讨论，见Israel（1987）。

7　Anderson（1929），Stoner（1930）。

8　图中所示白矮星质量和周长，以及钱德拉塞卡的白矮星内部结构的结果，后来发表在Chandrasekhar（1935）。

9　Eddington（1935a）。爱丁顿那些似乎有理的论证，见Eddington（1935b）。

10　Wali（1991）。

11　Wali（1991）。

12　这是我1958～1962年上大学时，加州理工学院的一个著名天文学教授用很权威的语气告诉我的。从那时起，我便强烈感到，大概多数天文学家自40年代初以来都有这种观点，也都在那么做，但我还是不能肯定。

13　引自Wali（1991）。

14　在批评这一章的初稿时，伊斯雷尔建议我这样解释爱丁顿的

行为；我相信这是符合历史事实的。

第5章

一般说明：本章的历史记述主要根据（i）我对所述事件的参与者和他们的科学家同事和朋友的访问（INT-Baym,INT-Braginsky,INT-Eggen,INT-Fowler,INT-Ginzbung,INT-Greenstein,INT-Harrison,INT-Khalatnikov,INT-Lifshitz,INT-Sandage,INT-Serber,INT-Volkoff,INT-Wheeler），和（ii）我所阅读的这些参与者写的文章。二三十年代物理学的历史背景，我多少依赖于Kevles（1971）；苏联物理学的历史背景，则根据Medvedev（1978）。关于朗道的背景资料，来自Livanova（1980），Gamow（1970）；关于奥本海默的，来自Rabi et al.（1969），Smith and Weiner（1980）；关于惠勒思想的发展，来自他的研究笔记，Wheeler（1988）；还有些地方则依据了下面引用的那些材料。

1　INT-Fowler。

2　INT-Greenstein和Greenstein（1982）。

3　Zwicky（1935）。

4　INT-Greenstein。

5　Baade（1952）。

6　这些数字是巴德和茨维基在图5.2复制的那个讲话摘要里公布的（Baade and Zwicky，1934a），不过“10 000或者也许1 000万”则出现在这个问题的一篇更详尽的文章里（Baade and Zwicky，1934b）。他们的误差来自假定在超新星最亮时，它的辐射热气体周长范围在1到100个太阳周长之间。实际上，周长远比这个大；如果我们追索他们的论证，会发现这个假定的紫外线和X射线的结果太小

了。

7　在这一节和整个这一章里，我都把中子星概念和它关于超新星和宇宙线的结果归功于茨维基，不过这些思想是他和巴德联合发表的。我相信，这些思想是茨维基的（我也相信巴德对观测数据的关键认识），根据来自与他们的科学同事的讨论：INT-Eggen,INT-Fowler,INT-Greenstein,INT-Sandage。

8　图5.2，Baade　and　Zwicky（1934a）。摘要里的数据，在Baade and Zwicky（1934b）有更详细的说明。

9　关于朗道发表这篇文章的原因，是他一生最亲密的伙伴栗弗席兹告诉我的（INT-Lifshitz）。

10　引自Livanova（1980）。

11　引自Livanova（1980）。

12　Gamow（1970）。

13　在斯大林时代入狱和死亡人数的统计还不太确切。Medveder（1978）提出的可能是70年代所能得到的最可靠数据。不过，80年代苏联开放以来，大众流传的消息又让这个数字大了好多。我引用的数是一个俄罗斯朋友的总体估计，他根据开放公布的材料对这一问题有过较深人的研究。

14　Eddington（1926）第11章及其参考文献。

15　Landau（1932）。

16　朗道的手稿发表在Landau（1938）。他不知道，他的亲密朋友盖莫夫已经发表过相同的思想（Gamow，1937）。1933年，斯

大林铁幕刚罩下不久，盖莫夫就逃离了苏联（见Gamow，1970）；但是，他在离开以前知道朗道关于致密中心核为恒星维持热量的原始中子核想法。中子发现以后，盖莫夫和朗道（这时两人已经失去了联系）自然会独立将朗道1931年的核解释为中子核。

17　1982年，朗道最亲密的朋友栗弗席兹提醒我注意那封信（INT-Lifshitz），并向我讲了它的背景，如文中所述。栗弗席兹去世后，这封信的全文——另外还有卡皮查与莫洛托夫、卡皮查与斯大林和卡皮查与别里亚的那些最终救朗道出狱的通信——发表在Khalatnikov（1988）。这里摘录的片段是我自己从俄文译过来的。

18　Gorelik（1991）。

19　见注17。

20　引自Royal（1969）。

21　Serber（1969）。

22　现在认为这些巨星是在双星系中形成的：一颗恒星坍缩成为中子星以后，经过很长时间，螺旋落进伴星的中心并在那儿留下来。这些怪物后来叫“Thorne-Żytkow天体”，因为乔特科夫（Anna Żytkow）和我最先详细计算了它们的结构。见Thorne and Żytkow（1977）；也见Cannon et al.（1992）。

23　Oppenheimer and Serber（1938）。

24　Shapiro and Teukolsky（1983）,Hartle and Sabbadini（1977）。

25　卡片5.4：我在这张卡片里对研究步骤的大多数描写都是猜测的，主要根据是对沃尔科夫的访问（INT-Volkoff），托尔曼档案

（Tolman,1948）和亲历者的文章（Oppenheimer and Volkoff,1939；Tolman,1939）。

26　托尔曼和奥本海默的通信见Tolman（1948）的档案。

27　INT-Volkoff.

28　这个结论发表在Oppenheimer and Volkoff（1939）。奥本海默和沃尔科夫对核力效应的估计所依据的托尔曼的解析分析，发表在Tolman（1939）。

29　Wheeler（1988）,Vol.4，p.33～40。

30　惠勒的背景和早期研究的详情见Wheeler（1979），Thorne and Zurek（1986）。

31　卡片5.5：这个物态方程（哈里森和惠勒的研究成果）发表在Harrison,Wakano,and Wheeler（1958），更详细的是Harrison,Thorne,Wakano,and Wheeler（1965）。近些年，Shapiro and Teukolsky（1983）的评述，核密度（$10^{14}$克/厘米$^3$）及其以上的实曲线是现代不同物态方程的近似。

32　图5.5：据Harrison,Wakano,and Wheeler（1958），和Harrison,Thorne,Wakano,and Wheeler（1965）。据Shapiro and Teukolsky（1983）的评述，中子星实曲线是现代不同计算的近似。

33　Oppenheimer and Volkoff（1939）。

34　Zwicky（1939）。

35　Rabi et al.（1969）。

第6章

一般说明：本章的历史记述主要根据（i）我对所说事件参与者和他们的科学家同事的访问（INT-Braginsky,INT-Finkelstein,INT-Fowler,INT-Ginzburg,INT-Harrison,INT-Lifshitz,INT-Misner,INT-Serber,INT-Wheeler,INT-Zel'dovich），（ii）我自己在某些事情的经历，（iii）参与者们写的科学论文，（iv）Bethe（1982），Rhodes（1986）5，Teller（1955）和York（1976）对美国核武器计划的记述，（v）Golovin（1973），Medvedev（1978），Ritus（1990），Romanov（1990）和Sakharov（1990）对苏联核武器和其他事件的记述，以及（vi）惠勒的研究笔记（Wheeler,1988）。

1　惠勒的演讲和他与奥本海默的交流，发表于Solvay（1958）。

2　这段话是Harrison,Wakano,and　Wheeler（1958）原话的大意，为适合本书表达习惯，文字上略有改动。

3　INT-Serber。

4　INT-Fowler。

5　INT-Serber。

6　这是我的猜测。我并不能肯定他这么快就完成了考察，但根据对奥本海默和他在研究完成后写的文章（Oppenheimer and Snyder,1939）内容的认识，我很相信他真那么做了。

7　奥本海默和斯尼德的研究结果发表在Oppenheimer and Snyder（1939）。

8　INT-Fowler。

9　INT-Lifshitz。

10　Wheeler（1979），这是惠勒对核物理研究的自传性记述。

11　Bohr and Wheeler（1939），Wheeler（1979）。玻尔和惠勒没有用他们在文章里的名称来称钚-239，但特纳（Louis A.Turner）直接根据他们的图4推测那是一种理想的持续链式反应的核，在一个著名的秘密备忘录中提出以这种燃料来做原子弹（Wheeler,1985）。

12　INT-Zel'dovich,Zel'dovch and Khariton（1939）。

13　关于惠勒所起的关键作用，见Klauder（1972），p.2～5。

14　引自奥本海默1945年10月16日在新墨西哥洛斯阿莫斯的一次讲话；见Good-child（1980），p.172。

15　Goodchild（1980），p.174。

16　Wheeler（1979）。

17　据《纽约时报》1993年1月14日A5版报道，哈里顿在莫斯科的一次演讲中公布了这些细节。

18　Medvedev（1979）。

19　1949年10月30日一般咨询委员会给美国原子能委员会的报告。见York（1976）的附录。

20　Bethe（1982）。

21　INT-Wheeler。

22　INT-Wheeler。

23　USAEC（1984），p.251。

24　INT-Wheeler。

25　苏联开始氢弹设计的日期似乎有些混乱，Sakharov（1990）定为1948年春，而Ginzburg（1990）定为1947年。

26　这是Sakharov（1990）确定的年月；Ginzburg（1990）定在1947年。

27　脚注：萨哈罗夫的猜测概况见Sakharov（1990）。泽尔多维奇的判断是口头告诉一位亲密的俄罗斯朋友的，他又转告了我。

28　金兹堡告诉我的，他当时在场。萨哈罗夫也在；据他回忆的英文本（Sakharov,1990），这句话是这样的："Our job is to kiss Zel' dovich's ass."（我们的工作是给泽尔多维奇舔屁股。）我对泽尔多维奇和萨哈罗夫复杂关系的个人看法，见Thorne（1991）。

29　朗道的这句话，有好几位苏联理论物理学家给我讲过。

30　Romanov（1990）。

31　不同原子弹爆炸的能量释放数据，我引自York（1976）。

32　Sakharov（1990）。

33　Romanov（1990），Sakharov（1990）。在一篇纪念萨哈罗夫的文章里，罗曼诺夫把这一发现归功于萨哈罗夫和泽尔多维奇。萨哈罗夫说，"我们理论部门的几个人大概同时有了这个想法，"但他接着说，"泽尔多维奇、特鲁特涅夫（Yuri Trutnev）和其他一些人无

疑做过重要贡献。”留给人们的印象是，他自己是最大的贡献者。

34　USAEC（1954）。

35　1991年7月，惠勒与索恩的电话交谈。

36　Sakharov（1990）。

37　Colgate and Johnson（1960）讨论了认识超新星和它作为宇宙线源的作用的动机。Colgate and White（1963,1966）用牛顿引力描述（而没用爱因斯坦的）进行了小质量超新星形成模拟。May and White（1965，1966）用爱因斯坦的引力的广义相对论描述做了大质量超新星形成的模拟。

38　Imshennik and Nadezhin（1964），Podurets（1964）。

39　INT-Lifshitz。

40　Finkelstein（1958）。

41　脚注：例如，可以看MTW卡片31.1和第31章的讨论。

42　这到底是怎么发现的，请看芬克尔斯坦自己的叙述，Finkelstein（1993）。

43　Thorne（1967）。

44　Harrison,Thorne,Wakano,and Wheeler（1965）。

45　Wheeler（1968）。

第7章

一般说明：本章的历史记述主要根据（i）我个人的亲身经历，（ii）我对其他参与者的访问（INT-Carter,INT-Chandrasekhar,INT-Detweiler,INT-Eardley,INT-Ellis,INT-Misner,INT-Novikov,INT-Penrose,INT-Press,INT-Price,INT-Rees,INT-Sciama,INT-Smarr,INT-Teukolsky,INT-Wald,INT-Wheeler,INT-Zel'dovich），和（iii）这些参与者写的科学论文。

1　Wheeler（1964b）。

2　我第一次发表环猜想是在纪念惠勒的一本文集里（Thorne，1972），也见MTW卡片32.3。

3　诺维科夫和泽尔多维奇称这个思想是半封闭宇宙。最后，他们单独发表了文章来讨论：Zel'dovich（1962），Novikov（1963）。

4　INT-Novikov。

5　INT-Novikov。

6　这一研究的关键思想和初始计算发表在Ginzburg（1964）；更完备的数学分析是金兹堡和一个年轻同事奥泽诺依（Leonid Moiseevich Ozernoy）完成的（Ginzburg and Ozernoy,1964）。

7　他们的分析和结果发表在Doroshkevich,Zel'dovich,and Novikoy（1965）（作者是照俄语字母顺序排的）。

8　会议前不久，诺维科夫和泽尔多维奇写过很有影响的评论；Zel'dovich and Novikov（1964,1965），读者可以从中体会诺维科夫的演说风格。

9　Doroshkevich,Zel'dovich,and Novikov（1965），见注7。

10　伊斯雷尔的分析发表在Israel（1967）。

11　Novikov（1969），de la Cruz,Chase,and Israel（1970）,Price（1972）。

12　de la Cruz,Chase,and Israel（1970）。

13　关于磁场与黑洞相互作用更详尽更完整的讨论，见Thorne,Price,and Macdonald（1986）图10，11和36。

14　评论和文献，见Carter（1979）6.7节；关于最后阶段的续篇发表在Mazur（1982）和Bunting（1983）。

15　Graves and Brill（1960）及其参考文献。

16　Kerr（1963）。

17　Carter（1966），Boyer and Lindquist（1967）。

18　Carter（1979）和其中的早期参考文献。

19　Carter（1968）。

20　Israel（1986）。

21　Penrose（1969）。

22　Newman et al.（1965）。

23　Press（1971）。

24　Teukolsky（1972）。

25　INT-Teukolsky。

26　Press and Teukolsky（1973）。

27　Chandrasekhar（1983b）。

第8章

一般说明：本章的历史记述主要根据（i）我个人的亲身经历，（ii）我对其他参与者的访问（INT-Giacconi,INT-Novikov,INT-Rees,INT-Van　Allen,INT-Zer'dovich），（iii）这些参与者写的科学论文，和（iv）下列发表的历史评述：Friedman（1972），Giacconi and Gursky（1974），Hirsh（1979）和Uhuru（1981）。

1　Wheeler（1964a）。

2　22年后的1986年，泽尔多维奇对我说，他很遗憾，当时在黑洞内部情况的问题上没能想得更远，INT-Zel'dovich。

3　Zel'dovich and Guseinov（1965）。

4　Trimble and Thorne（1969）。

5　Salpeter（1964），Zel'dovich（1964）。

6　Novikov and Zel'dovich（1966）。

7　Friedman（1972）。

8　Giacconi,Gursky,Paolini,and Rossi（1962）。

9　Sunyaev（1972）。

第9章

一般说明：本章的历史记述主要根据（i）我个人自1962年以来在这些事件的边缘的经历，（ii）我对其他参与者的访问（INT-Ginzburg,INT-Greenstein,INT-Rees,INT-Zel'dovich），（iii）那些参与者写的科学论文，和（iv）下列发表或没发表的历史记录：Hey（1973），Greenstein（1982），Kellermann and Sheets（1983），Struve and Ze-bergs（1962）和Sullivan（1982，1984）。

1　Jansky（1932）。

2　Whipple and Greenstein（1937）。

3　INT-Greenstein。

4　雷伯对自己工作的历史描述见Reber（1958）。

5　Reber（1940）。

6　INT-Greenstein。

7　INT-Greenstein。

8　Bolton,Stanley,and Slee（1949）。

9　Baade and Minkowski（1954）。

10　Jennison and Das Gupta（1953）。

11　这次会议的报告发表在Washington（1954）。

12　Schmidt（1963）。

13　Greenstein（1963）。

14　Smith（1965）。

15　Alfvén and Herlofson（1950），Kiepenheuer（1950），Ginzburg（1951）。这一工作历史的讨论，见Ginzburg（1984）。

16　Burbidge（1959）。

17　这个会议的报告发表在Robinson,Schild,and Shucking（1965）。

18　这是我根据自己对这次会议的回忆写的。

19　Rees（1971）。

20　Longair,Ryle,and Scheuer（1973）。

21　Salpeter（1964），Zel'dovich（1964）。

22　Lynden-Bell（1969）。

23　Bardeen and Petterson（1975）。

24　Bardeen（1970）。

25　Blandford and Rees（1974）。

26　Lynden-Bell（1978）。

27　Blandford（1976）。

28　Blandford and Znajek（1977）。

29　我们对类星体、射电星系、喷流和黑洞及其吸积盘作为它们的中心发动机的作用的认识现状，可以看Begelman,Blandford,and

Rees（1984），和Blandfond（1987），里面有详细论述。

30　例如，可以参考Phinney（1989）。

第10章

一般说明：本章的历史记述主要根据（i）我个人的亲身经历，（ii）我对其他参与者的访问（INT-Braginsky,INT-Drever,INT-Forward,INT-Grishchuk,INT-Weber,INT-Weiss），和（iii）那些参与者写的科学论文。对引力辐射和探测更专业的评述，可以参考Blair（1991）和Thorne（1987）。

1　Weber（1953）。

2　韦伯的工作成果发表在Weber（1960，1961）。6

3　1992年10月1日韦伯给我的信；那时他没有发表这个结果。韦伯的同事戴森第一个证明，大自然很可能在韦伯所选的频率附近产生引力波（Dyson，1963）。

4　Weber（1969）宣布观测到了引力波的证据。接下来的实验和是否真测到了引力波的争论，都记录在de Sabbata and Weber（1977）和它所引的论文里。这场争论的社会学研究，见Collins（1975，1981）。

5　暑期班的讲座，包括韦伯的，发表在DeWitt and DeWitt（1964）。

6　布拉金斯基警告的原话发表在Braginsky（1967）。

7　Braginsky（1977）和Giffard（1976）将警告说得更清楚了，而这个极限来自测不准原理的解释，在

Thorne,Drever,Caves,Zimmerman,and Sandberg（1978）。

8　1978年会议讨论的内容可以看Epstein and Clark（1979）。

9　Braginsky,Vorontsov,and Khalili（1978）；Thorne,Drever,Caves,Zimmermann,and Sandberg（1978）。

10　Michelson and Taber（1984）。

11　Gertsenshtein and Pustovoit（1962），Weber（1964），Weiss（1972），Mass,Miller,

and Forward（1971）。

12　例如，见Drever（1991）和它的参考文献。

13　见Braginsky and Khalili（1992）。

14　关于LIGO计划的回顾，见Abramovici et al.（1992）。

第11章

一般说明：本章（很少的一点）历史记述根据（i）我个人的经历，（ii）我对其他两个参与者的访问（INT-Damour,INT-Wald），（iii）他们写的科学论文，和（iv）我1965年在普林斯顿大学听过的库恩的科学革命和规范（范式）的课。

1　Kuhn（1962）。

2　本世纪最伟大的物理学家之一费曼（Richard Feynman）在一本可爱的小书《物理学定律的特征》（The Character of Physical Law,Feynman,1965）里，优美地描述了掌握几个规范会有多么强大的力量。不过，他从没用过“规范”这个词，我猜他没读过库恩的书。

库恩讲过人们多喜欢费曼的作风；费曼就是那样的。[7]

3　平直时空规范多少是由许多不同的人提出来的；专业上即大家知道的“广义相对论的平直时空体系的场论”。关于它的历史和概念的评述，见MTW以下章节：7.1，18.1；卡片7.1，17.2和18.1；练习7.3。它的优美推广解释了它与弯曲时空规范的关系，见Grishchuk,Petrov,and Popova（1984）。

4　Cohen and Wald（1971），Hanni and Ruffini（1973）。

5　Blandford and Znajek（1977）。

6　Znajek（1978），Damour（1978）。

7　Thorne,Price,and Macdonald（1986），也见Price and Thorne（1988）。

第12章

一般说明：本章的历史记述主要根据（i）我个人的经历，（ii）我对其他参与者的访问（INT-DeWitt,INT-Eardley,INT-Hartle,INT-Hawking,INT-Israel,INT-Penrose,INT-Unruh,INT-Wald,INT-Wheeler,INT-Zel'dovich），（iii）这些参与者写的科学论文，和（iv）下列发表的历史记录：Bekenstein（1980），Hawking（1988），Israel（1987）。

1　这里和后面关于霍金如何得到这个思想的记述，依据INT-Hawking和Hawking（1988）。[8]他的思想和具体结果发表在Hawking（1971b，1972，1973），本章第一节，“黑洞的生长”讲了大概内容。

2　Penrose（1963）.

3　卡片12.1：Hawking（1972，1973）。

4　INT-Israel,INT-Penrose,INT-Hawking。

5　Penrose（1965）。

6　卡片12.2：Hawking（1972，1973）。

7　Hawking and Hartle（1972）。

8　Christodoulou（1970）。

9　Bekenstein（1980）谈了这一点和后来与霍金的争论。贝肯斯坦关于黑洞熵的猜想和论证，发表在Bekenstein（1972,1973）。

10　1972年暑期班的报告发表在DeWitt and DeWitt（1973）。

11　Bardeen,Carter,and Hawking（1973）。

12　我1971年访问莫斯科时，米斯纳和惠勒也去了。但我和泽尔多维奇在他家讨论时，他们并不在场。

13　我凭记忆重构了下面的谈话，但说得不像我们讨论时那么专业。

14　Zel'dovich（1971）。

15　Zel'dovich and Starobinsky（1971）。

16　霍金在Hawking（1988）讲了他是如何“意外”发现黑洞辐射的。他的发现及其意义发表在Hawking（1974，1975，1976）。

17　例如，可以看Wald（1977）。

18　脚注：Wald（1977）。

19　Hawking（1988）。

20　Thorne,Price,and Macdonald（1986）第8章及其参考文献。

21　卡片12.5：Davies（1975），Unruh（1976）,Unruh and Wald（1982，1984）。

22　Gibbons and Hawking（1977）。

23　Page（1976）。

24　如，Hawking（1971a）；Novikov,Polnarev,Starobinsky,and Zel'dovich（1979）。

25　Page and Hawking（1975）；Novikov,Polnarev,Starobinsky,and Zel'dovich（1979）。

第13章

一般说明：本章的历史记述主要根据（i）我个人的经历（不过更像一个旁观者，而不是参与者），（ii）我对参与者的访问（INT-Belinsky,INT-DeWitt,INT-Geroch,INT-Khalatnikov,INT-Lifshitz,INT-MacCallum,INT-Misner,INT-Penrose,INT-Sciama,INT-Wheeler），和（iii）参与者们写的科学论文。

1　Harrison,Wakano,and Wheeler（1958）；Wheeler（1960）。

2　Wheeler（1964a,b）；Harrison,Thorne,Wakano,and Wheeler（1965）。

3　Oppenheimer and Snyder（1939）。

4　见第5章最后几段。

5　这里描写的奇点在坍缩恒星外的真空中；因为真空区域是用爱因斯坦方程的史瓦西解来描写的，所以这个奇点常指史瓦西几何奇点，MTW第32章有对它的定量分析。

6　图13.1：同上。

7　Wheeler（1960,1964a,b）；Harrison,Thorne,Wakano,and Wheeler（1965）。

8　卡拉特尼科夫和栗弗席兹得到这个结果的观点和计算发表在Lifshitz and Khalatnikov（1960，1963）和Landau and Lifshitz（1962）。

9　同上。

10　Landau and Lifshitz（1962）。

11　图13.4。Graves-Brill（1960）是在惠勒小组完成的，小组里的学生在60年代初觉得显然应该存在这里所说的那类爱因斯坦方程的解。然而，我和彭罗斯讨论发现，多数其他小组的研究者到60年代后期才知道这一点。这样的解很难具体构造，我们惠勒小组的人没有试过，也没发表过有关这个问题的东西。据我所知，这一思想和求解的努力，第一次发表在Navikov（1966）。

12　Graves and Brill（1960）及其参考文献。

13　关于彭罗斯的生平，主要根据INT-Penrose和INT-Sciama。

14　同上。

15　INT-Penrose,Penrose（1989）。

16　Penrose（1989）。

17　Penrose（1965）。

18　Hawking and Ellis（1973）的经典著作系统清理了整体方法。

19　Hawking and Penrose（1970）。

20　70年代我和栗弗席兹的私下讨论。

21　1990年6月18日卡拉特尼科夫给我的信。

22　这一段是我根据自己对会议和后来发生事情的回忆写的。

23　Khalatnikov and Lifshitz（1970），也见Belinsky,Khalatnikov,and Lifshitz（1970，1982）。

24　同上。

25　INT-Lifshitz,Livanova（1980）。

26　彭罗斯告诉我的。

27　Aleksandrov（1955,1959）。

28　Pimenov（1968）。

29　Lifshitz and Khalatnikov（1960，1963）。

30　例如，Novikov（1966）。

31　用专业的话说，不稳定的是Reissner-Nordström解的内柯西视界。这个猜想在Penrose（1968）；证明在Chandrasekhar and Hartle（1982）和它所引的早期文献。

32　Belinsky,Khalatnikov,and Lifshitz（1970，1982）。

33　Misner（1969）。

34　这是Wheeler（1960）根据他自己以前关于时空几何的真空涨落思想（Wheeler,1955，1957）第一次导出的结果。

35　脚注：Wheeler（1955，1957）引入了普朗克-惠勒时间并分析了它的物理意义。

36　这最早是由Wheeler（1960）提出的，后来通过现在所谓的“惠勒-德维特方程”而更定量化了。例如，可以参考Hawking（1987）的讨论。

37　Wheeler（1957，1960）。

38　例如，可以参考Hawking（1987，1988）。

39　Doroshkevich and Novikov（1978）证明奇点会衰退；Poisson and Israel（1990）和Ori（1991）用理想模型演绎了衰退的细节；Ori（1992）初步证明了这些模型是认识真实黑洞内奇点行为的很好指南。

40　模拟的具体情况见Shapiro and Teukolsky（1991）。

41　霍金的证据发表在Hawking（1992a）。

第14章

一般说明：这一章的历史记述几乎都是根据我个人的亲身经历。

1　Ludwig　Flamm（1916）发现，适当选择拓扑，爱因斯坦方程的Schwarzschild（1916a）解描写了空的球形虫洞。

2　图14.2：Kruskal（1960）。

3　Morris and Thorne（1988）。

4　Hawking and Ellis（1973）。

5　霍金只是根据他发现的黑洞蒸发很间接地推测了这一点，多少是试探性的。严格证明要等到6年以后的Candelas（1980）。

6　见Wald and Yurtsever（1991）和它引用的文献。

7　Wheeler（1955,1957，1960）。

8　Geroch（1967）。Friedman,Papastamatiou,Parker,and Zhang（1988）提出了格罗赫定理预言的虫洞生成的一个具体例子。

9　脚注：van　Stockum（1937），Gödel（1949），Tipler（1976）。

10　Morris,Thorne,and Yurtsever（1988）。

11　Morris,Thorne,and Yurtsever（1988）。

12　脚注：Friedman and Morris（1991）。

13　Echeverria,Klinkhammer,and Thorne（1991）。

14　卡片14.2：Echeverria,Klinkhammer,and

Thorne（1991）。

15　Forward（1992）。

16　以虫洞为时间机器引发的怪圈问题，Friedman et al（1990）有仔细而相当彻底的技术讨论。

17　Hall（1990）。

18　Hiscock and Konkowski（1982）。

19　Frolov（1991）。

20　Kim and Thorne（1991）。

21　Hawking（1992b）。

22　脚注：Gott（1991）。

23　Thorne（1993）多少从专业上讲了我怀疑时间机器的理由，并详细评述了到1993年春为止关于时间机器研究的状况。

[1]　即作者与米斯纳、惠勒所写的巨著《引力》（Gravitation），注释中的许多常识都引自此书。遗憾的是这部大作在大陆还没有汉译本，而现在流行的一些广义相对论著作都没有这些内容。照惯例，以下凡指示参考文献的英文名称都不再译成汉语。——译者

[2]　A.Pais这本书有中译本：上帝难以捉摸…——爱因斯坦的科学与生活，方在庆、李勇译，广东教育出版社，1998；爱因斯坦传，商务印书馆，2004。

[3]　这本书有中译本：相对论原理——狭义相对论和广义相对论经典论文集，赵志田，刘一贯译，孟昭英校，科学出版社，1980。——译者

[4]　中国曾出版过三卷本《爱因斯坦文集》（许良英、范岱年、赵中立等编译，商务印书馆，1979）；湖南科学技术出版社也正在组织翻译出版《爱因斯坦全集》，已出版5卷。——译者

[5] R.Rhodes这本书有中译本：《原子弹出世记》，李汇川、周文枚等译，世界知识出版社，1990。——译者

[6] Weber（1961）有汉译本：广义相对论与引力波，陈凤至、张大卫译，科学出版社，1977。

[7] 初学物理的读者看看三卷《费曼物理学讲义》（上海科学技术出版社），会得到别的任何课本都不会有的乐趣；费曼还有一本“玩笑的”自传，（《别逗了，费曼先生》，王祖哲译，湖南科学技术出版社，2005。）“看这本书而不笑的人，可能精神有问题。”——译者

[8] 即《时间简史》。——译者

# 文献目录

录音访问


Baym,Gordon.5 September 1985,Champaign/Urbana,Illinois.

Belinsky,Vladimir.27.March 1986.Moscow,U.S.S.R.

Braginsky,Vladimir Borisovich.20 December 1982,Moscow,U.S.S.R.;27 March 1986,Moscow,U.S.S.R.

Carter,Brandon.6 July 1983,Padova,Italy.

Chandrasekhar,Subrahmanyan.3 April 1982,Chicago,Illinois.

Damour,Thibault.26 July 1986,Cargese,Corsica.

Detweiler,Steven.December 1980,Baltimore,Maryland.

DeWitt,Bryce.December 1980,Baltimore,Maryland.

Drever,Ronald W.P.21 june 1982,Les Houches,France.

Eardley,Doug M.December 1980,Baltimore,Maryland.

Eggen,Olin.13 September 1985,Pasadena,California.

Ellis,George.December 1980,Baltimore,Maryland.

Finkelstein,David.8 July 1983,Padova,Italy.

Forward,Robert.31 August 1982,Oxnard,California.

Fowler,William A.6 August 1985,Pasadena,California.

Geroch,Robert.2 April 1982,Chicago,Illinois.

Giacconi,Riccardo.29 April 1983,Greenbelt,Maryland.

Ginzburg,Vitaly Lazarevich.December 1982,Moscow,U.S.S.R.;3 February 1989,Pasadena,California.

Greenstein,Jesse L.9 August 1985,Pasadena,California.

Grishchuk,Leonid P.26 March 1986,Moscow,U.S.S.R.

Harrison,B.Kent.5 September 1985,Provo,Utah.

Hartle,James B.December 1980,Baltimore,Maryland;2 April 1982,Chicago,Illinois.

Hawking,Stephen W.July 1980,Cambridge,England(not taped).

Ipaer,James R.December 1980,Baltimore,Maryland.

Israel,Werner.June 1982,Les Houches,France.

Khalatnikov,Isaac Markovich.27 March 1986,Moscow,U.S.S.R.

Lifshitz,Evgeny Michailovich.December 1982,Moscow,U.S.S.R.

MacCallum,Malcolm.30 August 1982,Santa Barbara,California.

Misner,Charles W.10 May 1981,Pasadena,California.

Novikov,Igor Dmitrievich.December 1982,Moscow,U.S.S.R.;28 March 1986,Moscow,U.S.S.R.

Penrose,Roger.7 July 1983,Padova,Italy.

Press,William H.December 1980,Baltimore,Maryland.

Price,Richard.December 1980,Baltimore,Maryland.

Rees,Martin.December 1980,Baltimore,Maryland.

Sandage,Allan.13 September 1985,Baltimore,Maryland.

Sciama,Dennis.8 July 1983,Padova,Italy.

Serber,Robert.5 August 1985,New York City.

Smarr,Larry.December 1980,Baltimore,Maryland.

Teukolsky,Saul A.27 January 1985,Ithaca,New York.

Unruh,William.December 1980,Baltimore,Maryland.

Van Allen,James.29 April 1973,Greenbelt,Maryland.

Volkoff,George.11 September 1985,Vancouver,British Columbia.

Wald,Robert M.December 1980,Baltimore,Maryland;2 April 1982,Chicago,Illinois.

Weber,Joseph.20 July 1982,College Park,Maryland.

Weiss,Rainer.7 July 1983,Padova,Italy.

Wheeler,John.December 1980,Baltimore,Maryland.

Zel' dovich,Yakov Borisovich.17 December 1982,Moscow,U.S.S.R.;22 and 27 March 1986,Moscow,U.S.S.R.

## 参考文献*

A bramovici,A.,Althouse,W.E.,Drever,R.W.P.,Gürsel,Y.,Kawamura,S.,Raab,F.J. Zucker,M.E.(1992).“LIGO:The Laser Interferometer Gravitational-W ave Observatory,”Science,256,325 -333.

Aleksandrov,A.D.(1955).“The Space-Time of the Theory of Relativity,”Helvetica Physica Acta,Supplement,4,4.Aleksandrov,A.D.(1959).“The Philosophical Implication and Significance of the Theory of Relativity,”Voprosy Filosofii,No.1,67.

Alfvén,H.,and Herlofson,N.(1950).“Cosmic Radiation and Radio Stars,”Physical Review,78,738.

Anderson,W.(1929).“Über die Grenzdichte der Materie und der Energie,”Zeitschrift für Physik,56,851.

Baade,W.(1952).“Report of the Commission on Extragalactic Nebulae,”Transactions of the International Astronomical U-nion,8,397.

Baade,W.,and Minkowski,R.(1954).“Identification of the Radio Sources in Cassiopeia,Cygnus A,and Puppis,”Astro-physical Journal,119,206.

Baade,W.,and Zwicky.F.(1934a).“Supernovae and Cosmic Rays,”Physical Review,45,138.

Baade,W.,and Zwicky,F.(1934b).“On Super-Novae,”Proceedings of the National Academy of Sciences,20,254.

Bardeen,J.M.(1970).“Kerr Metric Black Holes,”Nature,226,64.

Bardeen,J.M.,Carter,B.,and Hawking,S.W.(1973).“The Four Laws of Black Hole Mechanics,”Communications in Mathematical Physics,31,161.

Bardeen,J.M.,and Petterson,J.A.(1975).“The Lense -Thirring Effect and Accretion Disks around Kerr Black Holes,”Astrophysical Journal(Letters),195,L65 •

Begelman,M.C.,Blandford,R.D.,and Rees,M.J.(1984).“Theory of Extragalactic Radio Sources,”Reviews of Modern Physics,56,255.

Bekenstein.J.D.(1972).“Black Holes and the Second Law,”Lettere al Nuovo Cimento,4,131.

Bekenstein.J.D.(1973).“Black Holes and Entropy,”Physical Review D,7,2333.

Bekenstein,J.D.(1980).“Black Hole Thermodynamics,”Physics Today,January 24.

Belinsky,V.A.,Khalatnikov,I.M.,and Lifshitz,E.M.(1970).“Oscillatory Approach to a Singular Point in the Relativ-istic Cosmology,”Advances in Physics,19,525.

Belinsky,V.A.,Khalatnikov,I.M.,and Lifshitz,E.M.(1982).“Solution of the Einstein Equations with a Time Singularity,”Advances in Physics,31,639.

Bethe,H.A.(1982).“Comments on the History of the H-Bomb,”Los Alamos Science,Fall 1982,43.

Bethe,H.A.(1990).“Sakharov's H-Bomb,”Bulletin of the Atomic Scientists,October 1990.Reprinted in Drell and Kapitsa(1991),p.149.

Blair,D.,ed.(1991).The Detection of Gravitational Waves(Cambridge University Press,Cambridge,England).

Blandford,R.D.(1976).“Accretion Disc Electrodynamics—A Model for Double Radio Sources,”Monthly Notices of the Royal Astronomical Society,176,465.

Blandford,R.D.(1987).“Astrophysical Black Holes,”in 300 Years of Gravitation,edited by S.W.Hawking and W.Israel(Cambridge University Press,Cambridge,England),p.277.

Blandford,R.D.,and Rees,M.(1974).“A Twin-Exhaust Model for Double Radio Sources,”Monthly Notices of the Royal Astronomical Society,169,395.

Blandford,R.D.,and Znajek,R.L.(1977).“Electromagnetic Extraction of Energy from Kerr Black Holes,”Monthly Notices of the Royal Astronomical Society,179,433.

Hohr,N.,and Wheeler,J.A.(1939).“The Mechanism of Nuclear Fission,”Physical Review,56,426.

Moiton.J.G.,Stanley,G.J.,and Slee,0.B.(1949).“Positions of Three Discrete Sources of Galactic Radio-Frequency Radiation,”Nature,164,101.

Hoyer.R.H.,and Lindquist,R.W.(1967).“Maximal Analytic Extension of the Kerr Metric,”Journal of Mathematical Physics,8,265.

Hraginsky,V.B.(1967).“Classical and Quantum Restrictions on the Detection of Weak Disturbances of a Macroscopic Oscillator,”Zhurnal Eksperimentalnoi i Teoreticheskoi Fiziki,53,1434.English translation in Soviet Physics? — JETP,26,831(1968).

Braginsky,V.B.(1977).“The Detection of Gravitational Waves and Quantum Nondisturbtive Measurements,”in Topics in Theoretical and Experimental Gravitaton Physics,edited by V.de Sabbata and J.Weber(Plenum,London),p.105.

Braginsky,.V.B.,and Khalili,F.Ya.(1992).Quantum Measurements(Cambridge University Press,Cambridge,England).

Braginsky,V.B.,Vorontsov,Yu.I.,and Khalili,F.Ya.(1978).“Optimal Quantum Measurements in Detectors of Gravitational Radiation,”Pis' ma v Redaktsiyu Zhurnal Eksperimentalnoi i Teoreticheskoi Fiziki,27,296.English translation in JETP Letters,27,276(1978).

Braginsky,V,B.,Vorontsov,Yu.I.,and Thorne,K.S.(1980)“Quantum Nondemolition Measurements,”Science,209,547.

Brault,J.W.(1962).“The Gravitational Redshift in the Solar

Spectrum,”unpublished doctoral dissertation,Princeton University;available from University Microfilms,Ann Arbor,Michigan.

Brown,A.C.,ed.(1978).DROPSHOT:The American Plan for World War III against Russia in 1957(Dial Press/James Wade,New York).

Bunting,G.(1983).“Proof of the Uniqueness Conjecture for Black Holes,”unpublished Ph.D.dissertation,Department of Mathematics,University of New England,Armidale,N.S.W.Australia.

Burbidge,G.R.(1959).“The Theoretical Explanation of Radio Emission,”in Paris Symposium on Radio Astronomy,edited by R.N.Bracewell(Stanford University Press,Stanford,California).

Candelas,P.(1980).“Vacuum Polarization in Schwarzschild Spacetime,”Physical Review D,21,2185.

Cannon,R.C.,Eggleton,P.P.,Zytkow,A.N.,and Podsiadlowski,P,(1992).“The Structure and Evolution of Thorne—Zytkow Objects,”Astrophysical Journal,386,206—214.

Carter,B.(1966).“Complete Analytic Extension of the Symmetry Axis of Kerr's Solution of Einstein's Equations,”Physical Review,141,1242.

Carter,B.(1968).“Global Structure of the Kerr Family of Gravitational Fields,”Physical Review,174,1559.

Carter,B.(1979).“The General Theory of the Mechanical Electromagnetic and Thermodynamic Properties of Black Holes,”in

General Relativity:An Einstein Centenary Survey,edited by S.W.Hawking and W.Israel(Cambridge University Press,Cambridge,England),p.294.

Caves,C.M.,Thorne,K.S.,Drever,R.W.P.,Sandberg,V.D.,and Zimmermann,M.(1980).“On the Measurement of a Weak Classical Force Coupled to a Quantum-Mechanical Oscillator.I.Issues of Principle,”Reviews of Modem Physics,52,341.

Chandrasekhar,S.(1931).“The Maximum Mass of Ideal White Dwarfs,”Astrophysical Journal,74,81.

Chandrasekhar,S.(1935).“The Highly Collapsed Configurations of a Stellar Mass(Second Paper),”Monthly Notices of the Royal Astronomical Society,95,207.

Chandrasekhar,S.(1983a).Eddington:The Most Distinguished Astrophysicist of His Time(Cambridge University Press,Cambridge,England).

Chandrasekhar,S.(1983b).The Mathematical Theory of Black Holes(Oxford University Press,New York).

Chandrasekhar,S.(1989).Selected Papers of S.Chandrasekhar.Volume I:Stellar Structure and Stellar Atmospheres(University of Chicago Press,Chicago).

Chandrasekhar,S.,and Hartle,J.M.(1982).“On Crossing the Cauchy Horizon of a Reissner-Nordstroöm Black Hole,”Proceedings of the Royal Society of London,A384,301.

Christodoulou,D.(1970).“Reversible and Irreversible

Transformations in Black-Hole Physics,”Physical Review Letters,25,1596.

Clark,R.W.(1971).Einstein:The Life and Times(World Publishing Co.,New York).

Cohen,J.M.,and Wald,R.M.(1971).“Point Charge in the Vicinity of a Schwarzschild Black Hole,”Journal of Mathematical Physics,12,1845.

Colgate,S.A.,and Johnson,M.H.(I960).“Hydrodynamic Origin of Cosmic Rays,”Physical Review Letters,5,235.

Colgate,S.A.,and White,R.H.(1963).“Dynamics of a Supernova Explosion,”Bulletin of the American Physical Society,8,306.

Colgate,S.A.,and White,R.H.(1966).“The Hydrodynamic Behavior of Supernova Explosions,”Astrophysical Journal,143,626.

Collins,H.M.(1975).“The Seven Sexes:A Study in the Sociology of a Phenomenon,or the Replication of Experiments in Physics,”Sociology,9,205.

Collins,H.M.(1981).“Son of Seven Sexes:The Social Destruction of a Physical Phenomenon,”Social Studies of Science(SAGE,London and Beverly Hills),11,33.

Damour,T.(1978).“Black-Hole Eddy Currents,”Physical Review D,18,3598.

Davies,P.C.W.(1975).“Scalar Particle Production in Schwarzschild and Rindler Metrics,”Journal of Physics A,8,609.

de la Cruz,V.,Chase,J.E.,and Israel,W.(1970).“Gravitational Collapse with Asymmetries,”Physical Review Letters,24,423.

de Sabbata,V.,and Weber,J.,eds.(1977).Topics in Theoretical and Experimental Gravitation Physics(Plenum,New York).

DeWitt,C.,and DeWitt,B.S.,eds.(1964).Relativity,Groups,and Topology(Cordon and Breach,New York).

DeWitt,C.,and DeWitt,B.S.,eds.(1973).Black Holes(Gordon and Breach,New York).

Doroshkevich,A.D.,and Novikov,I.D.(1978).“Space-Time and Physical Fields in Black Holes,”Zhurnal Eksperi-mentalnoi i Teoreticheskii Fiziki,74,3.English translation in Soviet Physics—JETP,47,1(1978).

Doroshkevich,A.D.,Zel' dovich,Ya.B.,and Novikov,I.D.(1965).“Gravitational Collapse of Nonsymmetric and Rotating Masses,”Zhurnal Eksperimentalnoi i Teoreticheskii Fiziki,49,170.English translation in Soviet Physics—JETP,22,122(1966).

Drell,S.,and Kapitsa,S.,eds.(1991).Sakharov Remembered:A Tribute by Friends and Colleagues(American Institute of Physics,New York).

Drever,R.W.P.(1991).“Fabry-Perot Cavity Gravity-Wave Detectors,”in The Detection of Gravitational Waves,edited by D.Blair(Cambridge University Press,Cambridge,England),p.306.

Dyson,F.J.(1963).“Gravitational Machines,”in The Search for

Extraterrestrial Life,edited by A.G.W.Cameron(W.A.Benjamin,New York),p.115.

Echeverria,F.,Klinkhammer,G.,and Thorne,K.S.(1991).“Billiard Balls in Wormhole Spacetimes with Closed Timelike Curves.I.Classical Theory,”Physical Review D,44,1077.

ECP-1:Einstein,A.(1987).The Collected Papers of Albert Einstein.Volume 1:The Early Years,1879 -1902,edited by John Stachel(Princeton University Press,Princeton New Jersey).English translation by Anna Beck in a companion volume of the same title.

ECP-2:Einstein,A.(1989).The Collected Papers of Albert Einstein.Volume 2:The Swiss Years:Writings,1900,1909,edited by John Stachel(Princeton University Press,Princeton,New Jersey).English translation by Anna Beck in a companion volume of the same title.

Eddington,A.S.(1926).The Internal Constitution of the Stars(Cambridge University Press,Cambridge,England).

Eddington,A.S.(1935a).“Relativistic Degeneracy,”Observatory,58,37.

Eddington,A.S.(1935b).“On Relativistic Degeneracy,”Monthly Notices of the Royal Astronomical Society,95,194.

Einstein,A.(1911).“On the Influence of Gravity on the Propagation of Light,”Annalen der Physik,35,898.

Einstein.A.(1915).“The Field Equations for Gravitation,”Sitzungsberichte der Deutschen Akademie der

Wissenschaften zu Berlin,Klasse fur Mathematik,Physik,und Technik,1915,844.

Einstein,A.(1939)“On a Stationary System with Spherical Symmetry Consisting of Many Gravitating Masses,”Annals of Mathematics,40,922.

Einstein,A.(1949)“Autobiographical Notes,”in Albert Einstein:Philosopher-Scientist,edited by Paul A.Schilpp(Library of Living Philosophers,Evanston,Illinois).

Einstein,A.,and Mari¿,M.(1992).Albert Einstein/Mileva Mari¿:The Love Letters,edited by Jürgen Renn and Robert Schulman(Princeton University Press,Princeton,New Jersey).

Eisenstaedt,J.(1982).“Histoire et Singularités de la Solution de Schwarzschild,”Archive for History of Exact Sciences,27,157.

Eisenstaedt,J.(1991).“De l'Influence de la Gravitation sur la Propagation de la Lumiére en Théorie Newtonienne.L'Archéologie des Trous Noirs,”Archive for History of Exact Sciences,42,315.

Epstein,R.,and Clark,J.P.A.(1979).“Discussion Session II:Sources of Gravitational Radiation,”in Sources of Gravitational Radiation,edited by L.Smarr(Cambridge University Press,Cambridge,England),p.477.

Feynman,R.P.(1965).The Character of Physical Law(British Broadcasting Corporation,London;paperback edition:MIT Press,Cambridge,Massachusetts).

Finkelstein,D.(1958).“Past-Future Asymmetry of the

Gravitational Field of a Point Particle,"Physical Review,110,965.

Finkelstein,D.(1993)•"Misner,Kinks,and Black Holes,"in Directions in General Relativity.Volume 1:Papers in Honor of Charles Misner,edited by B.L.Hu,M.P.Ryan Jr.,and C.V.Vishveshwara(Cambridge University Press,Cambridge,England),p.99.

Flamm,L.(1916)."Beitrage zur Einsteinschen Cravitationstheorie,"Physik Zeitschrift,17,448.

Forward,R.L.(1992).Timemaster,(Tor Books,New York).

Fowler,R.H.(1926)•"On Dense Matter,"Monthly Notices of the Royal Astronomical Society,87,114.

Frank,P.(1947).Einstein:His life and Times(Alfred A.Knopf,New York).

Friedman,H.(1972)."Rocket Astronomy,"Annals of the New York Academy of Sciences,198,267.

Friedman,J.,and Morris,M.S.(1991)."The Cauchy Problem for the Scalar Wave Equation Is Well Defined on a Class of Spacetimes with Closed Timelike Curves,"Physical Review Letters,66,401.

Friedman,J.,Morris,M.S.,Novikov,I.D.,Echeverria,F.,Klinkhammer,G.,Th Yurtsever,U.

(1990)."Cauchy Problem in Spacetimes with Closed Timelike Curves,"Physical Review D,42,1915.

Friedman,J.,Papastamatiou,N.,Parker,L.,and Zhang,H.(1988).“Non-orientable Foam and an Effective Planck Mass for Point-like Fermions,”Nuclear Physics,B309,533;appendix.

Frolov,V.P.(1991).“Vacuum Polarization in a Locally Static Multiply Connected Spacetime and a Time-Machine Problem,”Physical Review D,43,3878.

Gamow,G.(1937).Structure of Atomic Nuclei and Nuclear Transformations(Clarendon Press,Oxford,England),pp.234-238.

Gamow,G.(1970).My World Line(Viking Press,New York).

Geroch,R.P.(1967).“Topology in General Relativity,”Journal of Mathematical Physics,8,782.

Gertsenshtein,M.E.,and Pustovoit,V.I.(1962).“On the Detection of Low-Fre quency Gravitational Waves,”Zhurnal Eksperimentalnoi i Teoreticheskoi Fiziki,43,605.English translation in Soviet Physics—JETP,16,433(1963).

Giacconi,R.,and Gursky,H.,eds(1974).X-Ray Astronomy(Reidel,Dordrecht,Holland).

Giacconi,R.,Gursky,H.,Paolini,F.R.,and Rossi,B.B.(1962).“Evidence for X-Rays from Sources Outside the Solar System,”Physical Review Letters,9,439.

Gibbons,G.(1979).“The Man Who Invented Black Holes,”New Scientist,28,1101(29 June).

Gibbons,G.W.,and Hawking,S.W.(1977).“Action Integrals and

Partition Functions in Quantum Gravity,"Physical Review D,15,2752.

Giffard,R.(1976)."Ultimate Sensitivity Limit of a Resonant Gravitational Wave Antenna Using a Linear Motion Detector,"Physical Review D,14,2478.

Ginzburg.V.L.,(1951)."Cosmic Rays as the Source of Galactic Radio Waves,"Doklady Akademii Nauk SSSR,76,377.

Ginzburg,V.L.(1964)."The Magnetic Fields of Collapsing Masses and the Nature of Superstars,"Doklady Akademii Nauk SSSR,156,43.English translation in Soviet Physics—Doklady,9,329(1964).

Cinzburg.V.L.(1984)."Some Remarks on the History of the Development of Radio Astronomy,"in The Early Years of Radio Astronomy,edited by W.J.Sullivan(Cambridge University Press,Cambridge,England).

Ginzburg,V.L.(1990).Private communication to K.S.Thorne.

Ginzburg,V.L.,and Ozemoy,L.M.(1964)."On Gravitational Collapse of Magnetic Stars,"Zhurnal Eksperimentalnoi i Teoreticheskoi Fiziki,47,1030.English translation in Soviet Physics—JETP,20,689(1965).

Gleick,J.(1987).Chaos:Making a New Science(Viking/Penguin,New York).

Gfidel,K.(1949)."An Example of a New Type of Cosmological Solution of Einstein's Field Equations of Gravitation,"Reviews of

Modern Physics,21,447.

Golovin,I.N.(1973).I.V.Kurchatov(Atomizdat,Moscow),2nd edition.An English translation of the earlier and less complete first edition was published as Academician Igor Kurchatov(Mir Publishers,Moscow,1969;also, Selbstverlag Press,Bloomington,Indiana,1968.)

Goodchild,P.(1980).J Robert Oppenheimer,Shatterer of Worlds(British Broadcasting Company,London).

Gorelik,G.E.(1991).“‘ My Anti-Soviet Activities…’ One Year in the Life of L.D.Landau,”Priroda,November issue,p.93;in Russian.

Gott,J.R.(1991).“Closed Timelike Curves Produced by Pairs of Moving Cosmic Strings:Exact Solutions,”Physical Review Letters,66,1126.

Graves,J.C.,and Brill,D.R.(1960).“Oscillitory Character of the Reissner-Nord-ström Metric for an Ideal Charged Wormhole,”Physical Review,120,1507.

Greensterin,J.L.(1963).“Red-shift of the Unusual Radio Source:3C48,”Nature,197,1041.

Greenstein.J.L.(1982).Oral history interview by Rachel Prud'homme,February and March 1982,Archives,California Institute of Technology.

Greenstein.J.L.,Oke,J.B.,and Shipman,H.(1985).“On the Redshift of Sirius B.”Quarterly Journal of the Royal Astronomical Society.26,279.

Grishchuk,L P.,Petrov,A.N.,and Popova,A.D.(1984).“Exact Theory of the Einstein Gravitational Field in an Arbitrary Background Space-Time,”Communications in Mathematical Physics,94,379.

Hall,S.S.(1989).“The Man Who Invented Time Travel:The Astounding World of Kip Thorne,”California,October,p.68.

Hanni,R.S.,and Ruffini,R.(1973).“Lines of Force of a Point Charge Near a Schwarzschild Black Hole,”Physical Review D,8,3259.

Harrison,B.K.,Thorne,K.S.,Wakano,M.,and Wheeler,J.A.(1965).Gravitation Theory and Gravitational Collapse(University of Chicago Press,Chicago).

Harrison.B.K.,Wakano,M.,and Wheeler,J.A.(1958)•“Matter-Energy at High Density:End Point of Thermonuclear Evolution.”in La Structure et l'Evolution de l'Univers,Onzième Conseil de Physique Solvay(Stoops,Brussels),p.124.

Hartle,J.B.,and Sabbadini,A.G.(1977).“The Equation of State and Bounds on the Mass of Nonrotating Neutron Stars,”Astrophysical Journal,213,831.

Hawking,S.W.(1971a).“Gravitationally Collapsed Objects of Very Low Mass,”Monthly Notices of the Royal Astronomical Society,152,75.

Hawking,S.W.(1971b).“Gravitational Radiation from Colliding Black Holes,”Physical Review Letters,26,1344.

Hawking,S.W.(1972).“Black Holes in General Relativity,”Communications in Mathematical Physics,25,152.

Hawking,S.W.(1973).“The Event Horizon,”in Black Holes,edited by C.DeWitt and B.S.DeWitt(Gordon and Breach,New York),p.1.

Hawking,S.W.(1974).“Black Hole Explosions?”Nature,248,30.

Hawking,S.W.(1975).“Particle Creation by Black Holes,”Communications in Mathematical Physics,43,199.

Hawking,S.W.(1976).“Black Holes and Thermodynamics,”Physical Review D,13,191.

Hawking,S.W.(1987).“Quantum Cosmology,”in 300 Years of Gravitation,edited by S.W.Hawking and W.Israel(Cambridge University Press,Cambridge,England),p.631.

Hawking,S.W.(1988).A Brief History of Time(Bantam Books,Toronto,New York).

Hawking,S.W.(1992a).“The Chronology Protection Conjecture,”Physical Review D,46,603.

Hawking,S.W.(1992b).“Evaporation of Two-Dimensional Black Holes,”Physical Review Letters,69,406.

Hawking,S.W.,and Ellis,G.F.R.(1973).The Large Scale Structure of Space-Time(Cambridge University Press,Cambridge,England).

Hawking,S.W.,and Hartle,J.B.(1972).“Energy and Angular

Momentum Flow into a Black Hole,”Communications in Mathematical Physics,27,283.

Hawking,S.W.,and Penrose,R.(1970).“The Singularities of Gravitational Collapse and Cosmology,”Proceedings of the Royal Society of London,A314,529.

Hey,J.S.(1973).The Evolution of Radio Astronomy(Neale Watson Academic Publications,Inc.,New York).

Hirsh,R.F.(1979).“Science,Technology,and Public Policy:The Case of X-Ray Astronomy,1959 to 1972,”unpublished Ph.D.dissertation,University of Wisconsin-Madison;available from University Microfilms,Ann Arbor,Michigan.

Hiscock,W.A.,and Konkowski,D.A.(1982).“Quantum Vacuum Energy in TaubNUT(Newman-Unti-Tamburi-no)-Type Cosmologies,”Physical Review D,6,1225.

Hoffman,B.(1972).In collaboration with H.Dukas,Albert Einstein:Creator and Rebel(Viking,New York).

Imshennik,V.S.,and Nadezhin,D.K.(1964).“Gas Dynamical Model of a Type II Supernova Outburst,”Astronomi-cheskii Zhurnal,41,829.English translation in Soviet Astronomy—AJ,8,664(1965).

Israel,W.(1967).“Event Horizons in Static Vacuum Spacetimes,”Physical Review,164,1776.

Israel,W.(1986).“Third Law of Black Hole Dynamics—A Formulation and Proof,”Physical Review Letters,57,397.

Israel,W.(1987).“Dark Stars:The Evolution of an Idea,”in 300 Years of Gravitation,edited by S.W.Hawking and W.Israel(Cambridge University Press,Cambridge,England),p.199.

Israel,W.(1990).Letter to K.S.Thorne,dated 28 May 1990,commenting on the semifinal draft of this book.

Jansky,K.(1932).“Directional Studies of Atmospherics at High Frequencies,”Proceedings of the Institute of Radio Engineers,20,1920.

Jennison,R.C.,and Das Gupta,M.K.(1953).“Fine Structure of the Extra-terrestrial Radio Source Cygnus 1,”Nature,172,996.

Kellermann,K.,and Sheets,B.(1983).Serendipitous Discoveries in Radio Astronomy(National Radio Astronomy Observatory,Green Bank,West Virginia).

Kerr,R.P.(1963).“Gravitational Field of a Spinning Mass as an Example of Algebraically Special Metrics,”Physical Review Letters,11,237.

Kevles,D.J.(1971).The Physicists(Random House,New York).

Khalatnikov,I.M.,ed.(1988).Vospominaniya o L.D.Landau(Nauka,Moscow).English translation;Landau,the Physicist and the Man:Recollections of L D.Landau(Pergamon Press,Oxford,England,1989).

Khalatnikov,I.M.,and Lifshitz,E.M.(1970).“The General Cosmological Solution of the Gravitational Equations with a Singularity in Time,”Physical Review Letters,24,76.

Kiepenheuer,K.O.(1950).“Cosmic Rays as the Source of General Galactic Radio Emission,”Physical Review,79,738.

Kim,S.-W.,and Thorne,K.S.(1991).“Do Vacuum Fluctuations Prevent the Creation of Closed Timelike Curves?”Physical Review D,43,3939.

Klauder,J.R.,ed.(1972).Magic without Magic:John Archibald Wheeler(W.H.Freeman,San Francisco).

Kruskal,M.D.(1960).“Minimal Extension of the Schwarzschild Metric,”Physical Review,119,1743.

Kuhn,T.(1962).The Structure of Scientific Revolutions(University of Chicago Press,chicago).

Landau,L.D.(1932).“On the Theory of Stars,”Physikalische Zeitschrift Sowjetunion,1,285.

Landau,L.D.(1938).“Origin of Stellar Energy,”Nature,141,333.

Landau,L.D.,and Lifshitz,E.M.(1962).Teoriya Polya(Gosudarstvennoye Izdatel'stvo Fiziko-Matematicheskoi Literaturi,Moscow),Section 108.English translation:The Classical Theory of Fields(Pergamon Press,Oxford,England,1962),Section 110.

Laplace,P.S.(1796).Exposition du Système du Monde.Volume II:Des Mouvements Réels des Corps Célestes(Paris).Published in English as The System of the World(W.Flint,London,1809).

Laplace,P.S.(1799).“Proof of the Theorem,that the Attractive

Force of a Heavenly Body Could Be so Large,that Light Could Not Flow Out of It,"Allgemeine Geographische Ephemeriden,verfasset von Einer Gesellschaft Gelehrten.8vo Weimer,IV,Bd I St.English translation in Appendix A of Hawking and Ellis(1973).

Lifshitz,E.M.,and Khalatnikov,I.M.(1960)."On the Singularities of Cosmological Solutions of the Gravitational E-quations.I."Zhurnal Eksperimentalnoi i Teoreticheskoi Fiziki,39,149.English translation in Soviet Physics— JETP,12,108 and 558(1961).

Lifshitz,E.M.,and Khalatnikov,I.M.(1963)."Investigations in Relativistic Cosmology,"Advances in Physics,12,185.

Livanova,A.(1980).Landau:A Great Physicist and Teacher(Pergamon Press,Oxford,England).

Longair,M.S.,Ryle,M.,and Scheuer,P.A.G.(1973)."Models of Extended Radio Sources,"Monthly Notices of the Royal Astronomical Society,164,243.

Lorentz,H.A.,Einstein,A.,Minkowski,H.,and Weyl,H.(1923).The Principle of Relativity;A Collection of Original Memoirs on the Special and General Theory of Relativity(Dover,New York).

Lynden-Bell,D.(1969),"Galactic Nuclei as Collapsed Old Quasars,"Nature,223,690.

Lynden-Bell,D.(1978)."Gravity Power,"Physica Scripta,17,185.

Mazur,P.(1982)."Proof of Uniqueness of the Kerr-Newman Black Hole Solution,"Journal of Physics A,15,3173.

May,M.ML.,and White,R.H,(1965).“Hydrodynamical Calculation of General Relativistic Collapse,”Bulletin of the American Physical Society,10,15.

May,M.M.,and White,R.H.(1966).“Hydrodynamic Calculations of General Relativistic Collapse,”Physical Review,141,1232.

Medvedev,Z.A.(1978).Soviet Science(W.W.Norton,New York).

Medvedev,Z.A.(1979).Nuclear Disaster in the Urals(W.W.Norton,New York).

Michell,J.(1784).“On the Means of Discovering the Distance,Magnitude,Etc.,of the Fixed Stars,in Consequence of the Diminution of Their Light,in Case Such a Diminution Should Be Found to Take Place in Any of Them,and Such Other Data Should Be Procured from Observations,as Would Be Further Necessary for That Purpose,”in Philosophical Transactions of the Royal Society of London,74,35;presented to the Royal Society on 27 November 1783.

Michelson,P.F.,and Taber,R.C.(1984).“Can a Resonant-Mass Gravitational-Wave Detector Have Wideband Sensitivity?”Physical Review D,29,2149.

Misner,C.W.(1969).“Mixmaster Universe,”Physical Review Letters,22,1071.

Misner,C.W.,Thorne,K.S.,and Wheeler,J.A.(1973).Gravitation(W.H.Freeman,San Francisco).

Mitton,S.,and Ryle,M.(1969).“High Resolution Observations of

Cygnus A at 2.7 GHz and 5 GHz,”Monthly Notices of the Royal Astronomical Society,146,221.

Morris,M.S.,and Thorne,K.S.(1988).“Wormholes in Spacetime and Their Use for Interstellar Travel:A Tool for Teaching General Relativity,”American Journal of Physics,56,395.

Morris,M.S.,Thorne,K.S.,and Yurtsever,U.(1988).“Wormholes,Time Machines,and the Weak Energy Condition,”Physical Review Letters,61,1446.

Moss,G.E.,Miller,L.R.,and Forward,R.L.,(1971).“Photon Noise Limited Laser Transducer for Gravitational Antenna,”Applied Optics,10,2495.

MTW:Misner,Thorne,and Wheeler(1973).

Newman,E.T.,Couch,E.,Chinnapared,K.,Exton,A.,Prakash,A.,and Torrence,R.(1965).“Metric of a Rotating,Charged Mass,”Journal of Mathematical Physics,6,918.

Novikov,I.D.(1963).“The Evolution of the Semi-Closed World,”Astronomicheskii Zhurnal,40,772.English translation in Soviet Astronomy—AJ,7,587(1964).

Novikov,I.D.(1966).“Change of Relativistic Collapse into Anticollapse and Kinematics of a Charged Sphere,”Pis'ma v Redaktsiyu Zhurnal Eksperimentalnoi i Teoreticheskoi Fiziki,3,223.English translation in JETP Letters,3,142(1966).

Novikov,I.D.(1969).“Metric Perturbations When Crossing the Schwarzschild Sphere,”Zhurnal Eksperimentlnoi i Teoreticheskoi

Fiziki,57,949.English translation in Soviet Physics—JETP,30,518(1970).

Novikov,I.D.,Polnarev,A.G.,Starobinsky,A.A.,and Zel'dovich,Ya.B.(1979).“Primordial Black Holes,”Astronomy and Astrophysics,80,104.

Novikov,I.D.,and Zel' dovich,Ya.B.(1966).“Physics of Relativistic Collapse,”Supplemento al Nuovo Cimento,4,810;Addendum 2.

Oppenheimer,J.R.,and Serber,R.(1938).“On the Stability of Stellar Neutron Cores,”Physical Review,54,608.

Oppenheimer,J.R.,and Snyder,H.(1939).“On Continued Gravitational Contraction.”Physical Review,56,455.

Oppenheimer,J.R.,and Volkoff,G.(1939).“On Massive Neutron Cores,”Physical Review,54,540.

Ori,A.(1991).“The Inner Structure of a Charged Black Hole:An Exact Mass Inflation Solution,”Physical Review Letters,67,789.

Ori,A.(1992).“Structure of the Singularity Inside a Realistic Rotating Black Hole,”Physical Review Letters,68,2117.

Page,D.N.(1976).“Particle Emission Rates from a Black Hole,”Physical Review D,13,198,and 14,3260.

Page,D.N.,and Hawking,S.W.(1975).“Gamma Rays from Primordial Black Holes,”Astrophysical Journal,206,1.

Pagels,H.(1982).The Cosmic Code(Simon and Schuster,New York).

Pais,A.(1982).“Subtle Is the Lord...”The Science and the Life of Albert Einstein(Oxford University Press,Oxford,England).

Penrose,R.(1965).“Gravitational Collapse and Spacetime Singularities,”Physical Review Letters,14,57.

Penrose,R.(1968).“The Structure of Spacetime,”in Battelle Rencontres:1967 Lectures in Mathematics and Physics,edited by C.M.DeWitt and J.A.Wheeler(Benjamin,New York),p.565.

Penrose,R.(1969).“Gravitational Collapse:The Role of General Relativity,”Rivista Nuovo Cimento,1,252.

Penrose,R.(1989).The Emperor's New Mind(Oxford University Press,New York),pp.419 -421.

Phinney,E.S.(1989).“Manifestations of a Massive Black Hole in the Galactic Center,”in The Center of the Galaxy:Proceedings of LAV Symposium 136 edited by M.Morris(Reidel,Dordrecht,Holland),p.543.

Pimenov,R.I.(1968).Prostranstva Kinimaticheskovo Tipa [ Seminars in Mathematics ],Vol.6(V.A.Steklov Mathematical Institute,Leningrad).English translation:Kinematic Spaces(Consultants Bureau,New York,1970).

Podurets,M.A.(1964).“The Collapse of a Star with Back Pressure Taken Into Account,”Doklady Akademi Nauk,154,300.English translation in Soviet Physics—

Doklady,9,1(1964).

Poisson,E.,and Israel,W.(1990).“Internal Structure of Black Holes,”Physical Review D,41,1796.

Press,W.H.(1971).“Long Wave Trains of Gravitational Waves from a Vibrating Black Hole,”Astrophysical Journal Letters,170,105.

Press,W.H.,and Teukolsky,S.A.(1973).“Perturbations of a Rotating Black Hole.II.Dynamical Stability of the Kerr Metric,”Astrophysical Journal,185,649.

Price,R.H.(1972).“Nonspherical Perturbations of Relativistic Gravitational Collapse,”Physical Review D,5,2419 and 2439.

Price,R.H.,and Thorne,K.S.(1988).“The Membrane Paradigm for Black Holes,”Scientific American,258(No.4),69.

Rabi,I.I.,Serber,R.,Weisskopf,V.F.,Pais,A.,and Seaborg,G.T.(1969).Oppenheimer(Scribners,New York).

Reber,G.(1940).“Cosmic Static,”Astrophysical Journal,91,621.

Reber,G.(1944).“Cosmic Static,”Astrophysical Journal,100,279.

Reber,G.(1958).“Early Radio Astronomy at Wheaton,Illinois,”Proceedings of the Institute of Radio Engineers,46,15.

Rees,M.(1971).“New Interpretation of Extragalactic Radio Sources,”Nature,229,312 and 510.

Renn,J.,and Schulman,R.(1992).Introduction to Albert

Einstein/Mileva Marić:The Love Letters,edited by Jürgen Renn and Robert Schulman(Princeton University Press,Princeton,New Jersey).

Rhodes,R.(1986).The Making of the Atomic Bomb(Simon and Schuster,New York).

Ritus V.I.(1990).“If Not I,Then Who?”Priroda,August issue.English translation in Drell and Kapitsa,eds.(1991).

Robinson,I.,Schild,A.,and Schucking,E.L.,eds.(1965).Quasi-Stellar Sources and Gravitational Collapse(University of Chicago Press,Chicago).

Romanov,Yu.A.(1990).“The Father of the Soviet Hydrogen Bomb,”Priroda,August issue.English translation in Drell and Kapitsa,eds.(1991).

Royal,D.(1969).The Story of J.Robert Oppenheimer(St.Martin's Press,New York).

Sagan,C.(1985).Contact(Simon and Schuster,New York).

Sakharov,A.(1990).Memoirs(Alfred A.Knopf,New York).

Salpeter,E.E.(1964).“Accretion of Interstellar Matter by Massive Objects,”Astrophysical Journal,140,796.

Schaffer,S.(1979).“John Michell and Black Holes,”Journal for the History of Astronomy,10,42.

Schmidt,M.(1963).“3C273:A Star-like Object with Large Red-shift,”Nature,197,1040.

Schwarzschild,K.(1916a).“Uber das Gravitationsfeld eines Massenpunktes nach der Einsteinschen Theorie,”Sitzungs-berichte der Deutschen Akademie der Wissenschaften zu Berlin,Klasse fur Mathematik,Physik,und Technik,1916,189.

Schwarzschild,K.(1916b).“Uber das Gravitationsfeld einer Kugel aus inkompressibler Flussigkeit nach der Einsteinschen Theorie,”Sitzungsberichte der Deutschen Akademie der Wissenschaften zu Berlin,Klasse fur Mathematik,Physik,und Technik,1916,424.

Seelig,C.(1956).Albert Einstein:A Documentary Biography(Staples Press,London),p.104.

Serber,R.(1969).“The Early Years,”in Rabi et al.(1969);also published in Physics Today,October 1967,p.35.

Shapiro,S.L.,and Teukolsky,S.A.(1983).Black Holes,White Dwarfs,and Neutron Stars(Wiley,New York).

Shapiro,S.L,and Teukolsky,S.A.(1991).“Formation of Naked Singularities—The Violation of Cosmic Censorship,”Physical Review Letters,66,994.

Smart,W.M.(1953).Celestial Mechanics(Longmans,Green and Co.,London),Section 19.03.

Smith,A.K.,and Weiner,C.(1980).Robert Oppenheimer.Letters and Recollections(Harvard University Press,Cambridge,Massachusetts).

Smith,H.J.(1965).“Light Variations of 3C273,”in Quasi-Stellar

Sources and Gravitational Collapse,edited by I.Robinson,A.Schild,and E.L.Schucking(University of Chicago Press,Chicago),p.221.

Solvay(1958).Onzième Conseil de Physique Solvay,La Structure et l'Evolution de l'Univers(Editions R.Stoops,Brussels).

Stoner,E.C.(1930).“The Equilibrium of Dense Stars,”Philosophical Magazine,9,944.

Struve,O.,and Zebergs,V.(1962).Astronomy of the 20th Century(Macmillan,New York).

Sullivan,W.J.,ed.(1982).Classics in Radio Astronomy(Reidel,Dordrecht,Holland).

Sullivan,W.J.,ed.(1984).The Early Years of Radio Astronomy(Cambridge University Press,Cambridge,England).

Sunyaev,R.A.(1972).“Variability of X Rays from Black Holes with Accretion Disks,”Astronomicheskii Zhurnal,49,1153.English translation in Soviet Astronomy—AJ,16,941,(1973).

Taylor,E.F.,and Wheeler,J.A.(1992).Spacetime Physics:Introduction to Special Relativity(W.H.Freeman,San Francisco).

Teller,E.(1955).“The Work of Many People,”Science,121,268.

Teukolsky,S.A.(1972).“Rotating Black Holes:Separable Wave Equations for Gravitational and Electromagnetic Perturbations,”Physical Review Letters,29,1115.

Thorne,K.S.(1967).“Gravitational Collapse,”Scientific American,217,(No.5),96.

Thorne,K.S.(1972).“Nonspherical Gravitational Collapse—A Short Review,”in Magic without Magic:John Archibald Wheeler,edited by J.R.Klauder(W.H.Freeman,San Francisco),p.231.

Thorne,K.S.(1974).“The Search for Black Holes,”Scientific American,231(No.6),32.

Thorne,K.S.(1987).“Gravitational Radiation,”in 300 Years of Gravitation,edited by S.W.Hawking and W.Israel(Cambridge University Press,Cambridge,England),p.330.

Thorne,K.S.(1991).“An American's Glimpses of Sakharov,”Priroda,May issue;in Russian.English translation in Drell and Kapitsa,eds.(1991),p.74.

Thorne,K.S.(1993).“Closed Timelike Curves,”in General Relativity and Gravitation 1992,edited by R.J.Gleiser,C.N.Kozameh,and D.M.Moreschi(Institute of Physics Publishing,Bristol,England),p.295.

Thorne,K.S.,Drever,R.W.P.,Caves,C.M.,Zimmermann,M.,and Sandberg,V.D.(1978).“Quantum Non-demolition Measurements of Harmonic Oscillators,”Physical Review Letters,40,667.

Thorne,K.S.,Price,R.H.,and Macdonald,D.A.,eds.(1986).Black Holes:The Membrane Paradigm(Yale University Press,New Haven,Connecticut).

Thorne,K.S.,and Zurek,W.(1986).“John Archibald Wheeler:A

Few Highlights of His Contributions to Physics,”Foundations of Physics,16,79.

Thorne,K.S.,and Zytkow,A.N.(1977).“Stars with Degenerate Neutron Cores.I.Structure of Equilibrium Models,”Astrophysical Journal,212,832.

Tipler,F.J.(1976).“Causality Violation in Asymptotically Flat Space-Times,”Physical Review Letters,37,879.

Tolman,R.C.(1939).“Static Solutions of Einstein's Field Equations for Spheres of Fluid,”Physical Review,55,364.

Tolman,R.C.(1948).The Richard Chace Tolman Papers,archived in the California Institute of Technology Archives.

Trimble,V.L.,and Thorne,K.S.(1969).“Spectroscopic Binaries and Collapsed Stars,”Astrophysical Journal,56,1013.

Uhuru(1981).“Proceedings of the Uhuru Memorial Symposium:The Past,Present,and Future of X-Ray Astronomy,”Journal of the Washington Academy of Sciences,71(No.1).

Unruh,W.G.(1976).“Notes on Black-Hole Evaporation,”Physical Review D,14,870.

Unruh,W.G.,and Wald,R,M.(1982).“Acceleration Radiation and the Generalized Second Law of Thermodynamics,”Physical Review D,25,942.

Unruh,W.D.,and Wald,R.M.(1984).“What Happens When an

Accelerating Observer Detects a Rindler Particle,"Physical Review D,29,1047.

USAEC [United States Atomic Energy Commission](1954).In the Matter of J.Robert Oppenheimer,Transcript of Hearing before Personnel Security Board,Washington,D.C.,April 12,1954,through May 6,1954(U.S.Government Printing Office,Washington,D.C.).

van Stockum,W.J.(1937)."The Gravitational Field of a Distribution of Particles Rotating about an Axis of Symmetry,"Proceedings of the Royal Society of Edinburgh,57,135.

Wald,R.M.(1977)."The Back Reaction Effect in Particle Creation in Curved Spacetime,"Communications in Mathematical Physics,54,1.

Wald,R.M.,and Yurtsever,U.(1991)."General Proof of the Averaged Null Energy Condition for a Massless Scalar Field in Two-Dimensional Curved Spacetime."Physical Review D,44,403.

Wali,K.C.(1991).Chandra:A Biography of S.Chandrasekhar(University of Chicago Press,Chicago).

Washington(1954)."Washington Conference on Radio Astronomy—1954,"Journal of Geophysical Research,59,1—204.

Weber,J.(1953)."Amplification of Microwave Radiation by Substances Not in Thermal Equilibrium,"Transactions of the IEEE,PG Electron Devices—3,1(June).

Weber,J.(1960)."Detection and Generation of Gravitational Waves,"Physical Review,117,306.

Weber,J.(1961).General Relativity and Gravitational Waves(Wiley-Interscience,New York).

Weber,J.(1964).Unpublished research notebooks;also documented in Robert Forward's unpublished Personal Journal No.C1338,page 66,13 September 1964.

Weber,J.(1969).“Evidence for Discovery of Gravitational Radiation,”Physical Review Letters,22,1320.

Weiss,R.(1972).“Electromagnetically Coupled Broadband Gravitational Antenna,”Quarterly Progress Report of the Research Laboratory of Electronics,M.I.T.,105,54.

Wheeler,J.A.(1955).“Geons,”Physical Review,97,511.Reprinted in Wheeler(1962),p.131.

Wheeler,J.A.(1957).“On the Nature of Quantum Geometrodynamics,”Annals of Physics,2,604.

Wheeler,J.A.(1960).“Neutrinos,Gravitation and Geometry,”in Proceedings of the International School of Physics,“Enrico Fermi,”Course XI(Zanichelli,Bologna).Reprinted in Wheeler(1962),p.1.

Wheeler,J.A.(1962).Geometrodynamics(Academic Press,New York).

Wheeler,J.A.(1964a).“The Superdense Star and the Critical Nucleon Number,”in Gravitation and Relativity,edited by H.Y.Chiu and W.F.Hoffman(Benjamin,New York),p.10.

Wheeler,J.A.(1964b).“Geometrodynamics and the Issue of the Final State,”in Relativity,Groups,and Topology,edited by C.DeWitt and B.S.DeWitt(Gordon and Breach,New York),p.315.

Wheeler,J.A.(1968).“Our Universe:The Known and the Unknown,”American Scientist,56,1.

Wheeler,J.A.(1979).“Some Men and Moments in the History of Nuclear Physics:The Interplay of Colleagues and Motivations,”in Nuclear Physics in Retrospect,edited by Roger H.Stuewer(University of Minnesota,Minneapolis).

Wheeler,J.A.(1985).Letter to K.S.Thorne dated 3 December.

Wheeler,J.A.(1988).Notebooks in which Wheeler recorded his research work and ideas as they developed;now archived at the American Philosophical Society Library,Philadelphia,Pennsylvania.

Wheeler,J.A.(1990).A Journey into Gravity and Spacetime(Scientific American Library,New York).

Whipple,F.L.,and Greenstein,J.L.(1937).“On the Origin of Interstellar Radio Disturbances,”Proceedings of the National Academy of Sciences,23,177.

White,T.H.(1939).The Once and Future King(Collins,London),Chapter 13 of Part I,“The Sword in the Stone.”

Will,C.M.(1986).Was Einstein Right?(Basic Books,New York).

York,H.(1976).The Advisors:Oppenheimer,Teller and the Superbome(W.H.Freeman,San Francisco).

Zel'dovich,Ya.B.(1962).“Semi-closed Worlds in the General Theory of Relativity,”Zhurnal Eksperimentalnoi i Teoreticheskoi Fiziki,43,1037.English translation in Soviet Physics—JETP,16,732(1963).

Zel'dovich,Ya.B.(1964).“The Fate of a Star and the Evolution of Gravitational Energy upon Accretion,”Doklady Akademii Nauk,155,67.English translation in Soviet Physics—Doklady,9,195(1964).

Zel'dovich,Ya.B.(1971).“The Generation of Waves by a Rotating Body,”Pis'ma v Redaktsiyu Zhurnal Eksperimentalnoi i Teoreticheskoi Fiziki,14,270.English translation in JETP Letters,14,180(1971).

Zel'dovich,Ya.B.(1985).Collected Works:Particles,Nuclei,and the Universe(Nauka,Moscow);in Russian.English translation:Selected Works of Yakov Borisovich Zel'dovich.Volume II:Particles,Nuclei,and the Universe(Princeton University Press,Princeton,1993).

Zel'dovich,Ya.B.,and Guseinov,0.Kh.(1965).“Collapsed Stars in Binaries,”Astrophysical Journal,144,840.

Zel'dovich,Ya.B.,and Khariton,Yu.B.(1939).“On the Issue of a Chain Reaction Based on an Isotope of Uranium.”Zhurnal Eksperimentalnoi i Teoreticheskoi Fiziki,9,1425;see also the follow-up papers by the same authors in the same journal,10,29(1940).and 10,477(1940).Reprinted as the first three papers in Volume II of Zel'dovich's collected works,Zel' dovich(1985).

Zel'dovich,Ya.B.,and Novikov,I.D.(1964).“Relativistic Astrophysics,Part I,”Uspekhi Fizicheskikh Nauk,84,877.English translation in Soviet Physics—Uspekhi,7,763(1965).

Zel'dovich,Ya.B.,and Novikov,I.D.(1965).“Relativistic Astrophysics,Part II,”Uspekhi FizicheskikhNauk,86,447.English translation in Soviet Physics—Uspekhi,8,522(1966).

Zel'dovich,Ya.B.and Starobinsky,A.A.(1971).“Particle Production and Vacuum Polarization in an Anisotropic Gravitational Field,”Zhurnal Eksperimentalnoi i Teoreticheskoi Fiziki,61,2161.English translation in Soviet Physics—JETP,34,1159(1972).

Znajek,R.(1978).“The Electric and Magnetic Conductivity of a Kerr Hole,”Monthly Notices of the Royal Astronomical Society,185,833.

Zwicky,F.(1935).“Stellar Guests,”Scientific Monthly,40,461.

Zwicky,F.(1939).“On the Theory and Observation of Highly Collapsed Stars,”Physical Review,55,726.

*关于黑洞研究更近而且系统的文献，请参阅1997年8月在德国Bad Honnef的一个暑期讲习班的讲义：Black Holes:Theory and Observation,Springer-Verlag,Berlin Heidelberg,1998.——译者

# 译后记

本书原来的副标题是Einstein's Outrageous Legacy，字中带韵，像一句诗；翻译过来大概是“爱因斯坦的奇异遗产”，就没那么好听了，而且读者对这个题目的联想，可能会离题太远；如果说“遗产”就是黑洞和时间弯曲，那又局限过多。

在确定译本题目的时候，我忽然想起《共产党宣言》开头那句很有名的话：“一个幽灵，共产主义的幽灵，在欧洲徘徊……”而在现代物理学中，我们处处能感觉到一个爱因斯坦的幽灵，那就是我们可以在这本书中看到的，为什么大智慧的物理学家们会去研究一些比小孩儿的问题还天真的东西，会“发明”和相信那些在普通人觉得荒谬的东西，会把数学的概念想象成宇宙中实际存在的东西。

我们在别的关于黑洞的书里，几乎只能看到藏在天上的奇迹，现在，索恩先生把更多的发生在物理学家头脑中的奇迹和故事端出来了，告诉大家，爱因斯坦的幽灵是如何在一个个物理问题上“出没”的，有时候物理学家却又借着爱翁的精神，把他那幽灵远远地丢在后面。

正如前言说的，这是一本历史，关于黑洞研究的历史，一部活的历史；几乎没有哪个问题有最后的答案，每一个有兴趣的读者，都可以走进来，甚至改写它。

时间机器原是幻想的东西，作者大概第一个把它认真当作物理学问题来研究。实际上，时间问题，在相对论、量子论里依然存在着，而且从本质上说，还是“经典的”，还不够革命；作者相信未来的量子

引力理论能够令人满意（霍金不久前说，量子论与相对论的结合很快就能实现），似乎也“不够革命”。我们现在还不知道量子引力以后的事情——从这点看，量子引力不过是我们面前一座突兀的高峰，它背后的峰谷不知还有多少！时间自古是哲学问题，当它成为物理学问题时，总会为物理学带来革命，我们今天的时间困惑又几乎要回归哲学了，当它再清晰地出现在物理学中时，我们大概会迎来新的物理学。

这本书原来请湖南师范大学的朱久运和黄亦斌先生译过，两家的语言风格相距较远，一时难得统一起来，译者只好重译，译得匆忙，没能采纳两家的成果，很遗憾。

虽然这是一本科普读物，但作者像写专著那样写，在重要的问题上，差不多“无一字无来处”；不过另一方面，像大多数科普读物一样，读者会看到许多重复的东西，也会遇到一些模糊的东西。这也是科学旅行的乐趣，不但能常遇老朋友，也会邂逅陌生人，虽然老朋友爱唠叨，陌生人又走得太匆匆，但一路上总不会寂寞。读者可能对语言环境感到陌生的东西，译者注里提供了一点信息，可能会有些帮助；关于物理学的东西，作者提到了一些很有影响的著作，译者见过中译本的，都特别说明了，请读者在那些书里去熟悉某些陌生的朋友。

我写最后这几行字也正在经历一种时间旅行，从旧千年走进新千年——当大家说“千禧之年”时，是不是想过，该有好多灵魂“复活”？（《新约·启示录》）那么，让我们祈祷：复活吧，爱因斯坦的幽灵……

译者

2000年元旦，成都

## 在黑洞的地平线上——重印后记

南山的雪闪着耀眼的光芒，勃朗峰直插我们头上的天空；在我们周围，牛群带着铃响在绿油油的牧场上吃草；山下离学校几百米的地方，是美丽如画的莱苏什的村庄……

大多数下午的时间我们都在不断讨论新的问题：诺维科夫和我关在小木屋里，想发现吸积到黑洞的气体是如何发射X射线的；在学校休息室的长椅上，我的学生普雷斯和特奥科尔斯基在讨论小干扰下的黑洞是不是还稳定；在50米外的山坡上，巴丁、卡特尔和霍金在全神贯注地用爱因斯坦的广义相对论推导完整的黑洞演化方程组。那真是难忘的田园诗，醉人的物理学！

我十多年前学相对论时，把图书馆里大大小小的相对论著作都找来看了。当然也有些科普的，但除了几个书名今天都记不起来了。为什么？因为它们不过是一些没有数学的教科书。学过数学以后，就遗憾地过河拆桥了。这样的科普读物很多，它们像风景区的路标，将陌生的游客引向一座座险峰，却几乎永远不能伴着风光走进人们美好的回忆。幸运的是，在《黑洞与时间弯曲》里的这幅60年代田园物理学风情画，自然令我想起海森伯对20年代的回忆，那是他发现量子力学矩阵形式的那些天，哦，多么壮丽的发现！

那是1925年5月底，我患了严重的花粉热，只好向玻恩请了半个月的假，直接去了赫里戈兰，我希望在远离花草的海滨，那令人心旷神怡的空气能很快让我恢复健康……我的房间在三楼，能看到村庄远处的沙滩和大海的壮丽景象。我坐在阳台上，反复地考虑着玻尔说过的话。

……结果，差不多到凌晨三点，我才最后算完……我太兴奋了，通宵未睡。黎明时，我朝岛的南端走去。我曾渴望登上那块伸向大海的岩石。没费多大力气我就爬上去了，在那儿等着日出。（海森伯《物理学及其他》）

我读这段话时，不知道量子理论是什么；今天，虽然我大概懂得了它的数学（照狄拉克的说法是一套计算法则），还是不知道它到底说了什么。不过，在量子论的发现的历史中，我经历了许多令人向往的思想奇迹。至今还影响着我的一句话，也在海森伯的这段回忆里，那是爱因斯坦对他说的："在原则上，单靠可观测量去建立一个理论，是完全错误的。实际上，正好相反，是理论决定我们观测到什么。"这差不多是我所理解的"爱因斯坦的幽灵"。在黑洞发现的经过里，这句话有着更加生动的表现，大概也更能够唤醒读者也许因为生活太累而昏睡了的科学理想。

一条历史的河流，当然比"过了河的桥"更值得人们回忆。在我看来，索恩的《黑洞与时间弯曲》，首先是一段活的历史，爱因斯坦身后的相对论历史。关于这段历史，我没见过什么系统的读物。我想，一方面，它涉及的数学太多，离人们生活太远，不会引起大众的关心；另一方面，它确定的东西太少，离我们时代太近，很难形成专门的话题。也正因为这些，读者才是幸运的。正如策划者之一的Frederick Seitz博士在《前言》里说的，"读这本书的人应怀着两个目标：学一些我们物理宇宙中的尽管奇异却很真实的可靠事实；欣赏那些我们还不那么有把握的奇思妙想。"当然，在别的关于"黑洞"的读物里，我们也能看到这些；不过，作者写得更真切，因为他是那些事实的经历者，是某些奇思妙想的参与者。"我和我要讲的东西关系太近了，我个人从60年代到今天都在亲历它的发展，我最好的几个朋友从30年代起就身在其中了。"我们看一个思想产生的经过，其实要比看它的结果有趣得多。

最显著的例子就是作者关于时间机器的叙述。多年来，科幻小说

和一般的科普读物给很多读者留下了不太正确的印象。实际上，一定物理条件下（如弱能量条件）产生的一定的数学结构（如类时闭曲线），只能在物理学概念的基础上讨论，借一个纯数学结果来展开想象，是没有什么意义的。反过来说，我们读一本内容陌生的书，谈一个没有最后答案的问题，最重要的还是思想和问题的过程，而不是结果——也许很久都不会有结果。读者从这些疑惑中，或许会萌发一点雄心，与作者开宗明义表达的心愿产生某种共鸣："30年来，我一直在探索，为的是去认识爱因斯坦为后代留下的遗产……去发现相对论失败的地方，看它如何失败，会有什么来取代它。"这其实也正是爱因斯坦向往过的牛顿的幸运："幸运啊，牛顿；幸福啊，科学的童年！"（为牛顿《光学》写的序），像朗道那样抱怨自己生得太晚（第5章），没有赶上百年前物理学革命年代的人，应该幸运地感到在新世纪的门口，遭遇了令人更加困惑的问题，从某种意义上说，主题还是当年爱因斯坦与玻尔的对话；在这一点上，我们离爱因斯坦的幽灵并不遥远。

这本书，从爱因斯坦时代走到90年代，把遗产和火炬接过来，又传下去；只有一个幽灵不变地在读者眼前徘徊，那个永远追求统一的幽灵，今天落在广义相对论和量子论之间，落在黑洞的地平线上……

译者

2000年8月

# 重印后记

黑洞的书近年来更多了，黑洞的理论也有了一些新认识（例如在宇宙全息观点下的认识）。但本书的基本内容并没有过时，2000年6月作者在60岁生日的纪念文集里，还津津乐道他和夫人的那个时间机器。当然，本书更持久的价值在于它在科学大背景下叙述了黑洞物理学的成长。

借重新包装的机会，我通读了全书，除修改一些字句，更多的倒是在重新经历那个“黄金年代”。在物理学的历史中，量子力学的诞生，曾经是最令人激动的一幕大戏——相对论尽管动人，却几乎是爱因斯坦个人的独角戏；经典的物理学虽然辉煌，但缺乏现代物理学的“纯粹理性”的趣味——而黑洞的历史（当然也许还有超弦的历史），现在看来更加有趣，因为它的数学和物理还正在进行着，还没进入普通大学生的课堂，更没尘封在科学史家的档案柜。读者读它不仅温故，而且知新，借孔夫子的话说，“可以为师矣！”

对本书的新读者，我推荐一种新的读法。你可以把它分解成三本：一本传奇，讲物理学家的故事；一本科普，讲黑洞和相关现象的物理学常识；一本手册，讲“我凭什么相信我说的”——正文之外的注释和文献，是黑洞物理学的历史和理论的良好导引。

多数读者，特别喜欢科学的同学，应该多读“第一本”，它可能是最有启发的。我们习惯了在欣赏文艺作品时联想它的作者和风格。所谓“文如其人”，说的是一种境界，作者的境界和表现在作品的境界。遗憾的是，人们似乎忘了科学家也是人，也有风格。风格决定了他选择的问题，而问题决定了科学发展的方向。常有人问，假如没有爱因

斯坦，会出现相对论吗？问题本身没有意义了，但相对论与量子物理学那么不同，我们似乎真可以说，那是因为它们本是不同父母的孩子。想进科学大门的同学，应该先学会科学家怎么做人，怎么做研究。本书就讲了好多大师的故事，特别讲了他们的风格。例如，黑洞的三个导师，“都有自己的风格。事实上，恐怕难以找到比这更鲜明的风格了。”惠勒是幻想家，泽尔多维奇像火种，而席艾玛像蜡烛。当然还有别的个性和风格，如奥本海默的小心谨慎，茨维基的大胆猜想，这些在物理学的“正史”里大概都是看不到的。科学成果的形式（如论文和报告）几乎掩盖了科学家的个性，而他们的个性有时更迷人。圈子里的人也许知道，学生选择老师，不一定看他的成果，也看他的“人”。套一句老话说，科学是没有人情味的，但科学家不能没有风格。同学们在这儿能遇到那么多物理学的导师，随便你喜欢哪个，跟着他走，也许能走出自己的新天地。

本书的尾声还在继续。作者发起的LIGO计划，自1999年运行以来，发现了不少引力波的信息。一本讲LIGO故事的书（《爱因斯坦的未完成交响曲》）也即将翻译出版了。LIGO计划还向中学生开放。正如LIGO实验室的DeSalvo博士说的，通过科学家与中学生的合作，“能吸引天才少年走近科学和技术，远离不那么迷人的商务和法律之类的东西……也就是那些靠脑子赚钱的领域。”更有趣的是，在2005年国际物理年时，美国物理学会支持了一个“爱因斯坦在家”（Einstein@Home）的计划，任何人都可以加入进来，在他个人电脑的“闲暇”时间，利用LIGO和GEO的数据寻找引力波源（如脉冲星，即旋转的中子星）。分析的时候，还能从屏幕保护程序看到你正在搜寻的那块天空呢。当LIGO胜利的那一天，也许有读者能自豪地说，“那胜利也有我的一份功劳呢！”

译者

2006年4月23日，世界读书日

# 宇宙的琴弦

# 总序

科学，特别是自然科学，最重要的目标之一，就是追寻科学本身的原动力，或曰追寻其第一推动。同时，科学的这种追求精神本身，又成为社会发展和人类进步的一种最基本的推动。

科学总是寻求发现和了解客观世界的新现象，研究和掌握新规律，总是在不懈地追求真理。科学是认真的、严谨的、实事求是的，同时，科学又是创造的。科学的最基本态度之一就是疑问，科学的最基本精神之一就是批判。

的确，科学活动，特别是自然科学活动，比较起其他的人类活动来，其最基本特征就是不断进步。哪怕在其他方面倒退的时候，科学却总是进步着，即使是缓慢而艰难地进步，这表明，自然科学活动中包含着人类的最进步因素。

正是在这个意义上，科学堪称为人类进步的“第一推动”。

科学教育，特别是自然科学的教育，是提高人们素质的重要因素，是现代教育的一个核心。科学教育不仅使人获得生活和工作所需的知识和技能，更重要的是使人获得科学思想、科学精神、科学态度以及科学方法的熏陶和培养，使人获得非生物本能的智慧，获得非与生俱来的灵魂。可以这样说，没有科学的“教育”，只是培养信仰，而不是教育。没有受过科学教育的人，只能称为受过训练，而非受过教育。

正是在这个意义上，科学堪称为使人进化为现代人的“第一推

动”。

近百年来，无数仁人智士意识到，强国富民再造中国离不开科学技术，他们为摆脱愚昧与无知做了艰苦卓绝的奋斗，中国的科学先贤们代代相传，不遗余力地为中国的进步献身于科学启蒙运动，以图完成国人的强国梦。然而应该说，这个目标远未达到。今日的中国需要新的科学启蒙，需要现代科学教育。只有全社会的人具备较高的科学素质，以科学的精神和思想、科学的态度和方法作为探讨和解决各类问题的共同基础和出发点，社会才能更好地向前发展和进步。因此，中国的进步离不开科学，是毋庸置疑的。

正是在这个意义上，似乎可以说，科学已被公认是中国进步所必不可少的推动。

然而，这并不意味着，科学的精神也同样地被公认和接受。虽然，科学已渗透到社会的各个领域和层面，科学的价值和地位也更高了，但是，毋庸讳言，在一定的范围内，或某些特定时候，人们只是承认“科学是有用的”，只停留在对科学所带来的后果的接受和承认，而不是对科学的原动力、科学的精神的接受和承认。此种现象的存在也是不能忽视的。

科学的精神之一，是它自身就是自身的“第一推动”。也就是说，科学活动在原则上是不隶属于服务于神学的，不隶属于服务于儒学的，科学活动在原则上也不隶属于服务于任何哲学。科学是超越宗教差别的，超越民族差别的，超越党派差别的，超越文化和地域的差别的，科学是普适的、独立的，它自身就是自身的主宰。

湖南科学技术出版社精选了一批关于科学思想和科学精神的世界名著，请有关学者译成中文出版，其目的就是为了传播科学的精神，科学的思想，特别是自然科学的精神和思想，从而起到倡导科学精神，推动科技发展，对全民进行新的科学启蒙和科学教育的作用，为

中国的进步做一点推动。丛书定名为《第一推动》，当然并非说其中每一册都是第一推动，但是可以肯定，蕴含在每一册中的科学的内容、观点、思想和精神，都会使你或多或少地更接近第一推动，或多或少地发现，自身如何成为自身的主宰。

《第一推动丛书》编委会

# 序

爱因斯坦在生命的最后30年里一直在寻找所谓的统一场论——一个能在单独的包罗万象的协和的框架下描绘自然力的理论。激励爱因斯坦的不是我们常想的那些与科学事业相关的东西，例如，为了解释这样或那样的实验数据。实际上，驱使他的是一种热忱的信念：对宇宙的最深刻认识将揭示一个最大的宇宙奇迹，那就是，它所依赖的基本原理是那么简单而有力。爱因斯坦渴望以前所未有的清晰来表现宇宙的活动，让每一个人都敬畏它那美妙动人的旋律。

爱因斯坦从未实现他的梦，主要原因是那底牌还没看清楚：那时，自然力和物质的许多基本特征我们还不知道，或者知之甚少。但在过去的半个世纪，每一代新生物理学家——经历无数的曲折，走过数不清的死胡同——都不断在前辈的基础上添砖加瓦，构筑起越来越完整的宇宙行为知识体系。当年，爱因斯坦满怀热情追求统一理论，却空手归来；如今，物理学家相信他们终于发现了一个框架，能把这些知识缝合成一个无缝的整体——一个单一的理论，一个原则上能描述一切现象的理论，这就是超弦理论，我们这本书的主题。

我写《宇宙的琴弦》，是为了把物理学研究前沿的惊人发现带给广大的读者，特别是那些没有经过数学和物理学训练的人。在过去的几年里，我开过一些超弦理论的普及演讲，发现很多人都渴望了解当代研究说了哪些关于宇宙定律的东西，那些定律如何要求重建一个不朽的宇宙概念，在对终极理论追求的背后，藏着哪些挑战。这本书解释了爱因斯坦、海森堡以来的主要物理学成就，描述了那些发现是如何在我们时代的科学突破中四处开花结果的，我希望这能丰富读者的

知识和满足读者的好奇。

我也希望《宇宙的琴弦》能令那些有一定科学修养的读者感兴趣，对自然科学的学生和老师来说，我希望这本书不但能具体提供一些现代物理学的基本材料，如狭义相对论、广义相对论和量子力学；同时也能把从四面八方走来寻求统一理论的研究者们动人的兴奋和激动传给大家。对于热心的科普读者，我向他们解释了近10年来我们在认识宇宙的过程中获得的振奋人心的进展；对于其他学科领域的同事，我希望这本书能给他们一种忠实而平静的感觉，使人了解为什么弦理论家会那么津津乐道追求终极自然理论的那么一点点进步。

超弦理论撒开了一张大网。它是一个深广的主题，融合着许多重要的物理学发现。这个理论统一了大与小的定律，大到统领宇宙的尽头，小到深入物质的核心。我们能通过许多不同的道路走近它。我选择的是我们不断演化着的空间和时间的认识，我认为这是一条特别扣人心弦的发展道路，它扫荡旧观念，引来了许多迷人的新发现。爱因斯坦向世界证明空间和时间在以一种陌生的令人惊讶的方式活动着。如今，前沿的研究已经通过许多卷缩在宇宙纤维里的隐藏维度把他的发现综合进量子宇宙——那些维度的复杂几何很可能是打开某些空前幽深的问题的钥匙。我们将看到，尽管有些概念令人难以捉摸，还是可以通过实际的类比来把握它们。理解了这些思想，一个惊人的革命性的宇宙图景将展现在面前。

贯穿全书，我都紧扣科学，同时也常常通过类比和比喻，让读者对科学家如何达到当今的宇宙概念有一个直观的认识。尽管我避开了专业术语和数学方程，但因为涉及的新概念太多，为了能完全跟上概念的发展，读者可能还得不时停下来，想想这儿，想想那儿。第四篇的几章（集中谈最新进展）比其他部分更抽象；我会小心地先警告读者，内容结构也经过了适当安排，以便读者可以匆匆浏览或者跳过它们，而不会对全书的逻辑有太大的影响。为便于读者记住正文里引进的概念，我编了一个科学词汇表。当然，马虎的读者可能会完全跳过

书后的注释，但认真的读者会在注释中看到正文的一些观点被扩充了，简化的思想也更清晰了，经过数学训练的人还能在那儿发现一些更富专业情趣的东西。

我在写这本书的过程中得到过许多人的帮助，我要感谢他们。David Steinhardt以极大的耐心阅读了原稿，以编辑的眼光慷慨地提出了很好的建议，并给了我极大的鼓励。David Morrison, Ken Vineberg, Raphael Kasper, Nicholas Boles, Steven Carlip, Arthur Greenspoon, David Mermin, Michael Popowits和Shani Offen认真读了原稿，具体谈了读后的感觉，提出了令本书大为增色的意见。另外还有不少朋友也读过全部原稿，并提出了建议和鼓励，他们是Paul Aspinwall, Persis Drell, Michael Duff, Kurt Gottfried, Joshua Greene, Teddy Jefferson, Marc Kamionkowski, Yakov Kanter, Andras Kovacs, David Lee, Megan McEwen, Nari Mistry, Hasan Padamsee, Ronen Plesser, Massimo Poratti, Fred Sherry, Lars Straeter, Steven Strogatz, Andrew Strominger, Henry Tye, Cumrun Vafa和Gabriele Veneziano。我要特别感谢Raphael Gunner，他在本书写作初期曾提出过很有远见的批评，使它在整体上能有现在的样子；我还要特别感谢Robert Malley，从本书的思考到落笔，他总是在一直激励着我。另外，Steven Weinberg和Sidney Coleman也给了我重要的指导和帮助。我很高兴与下列朋友进行过有益的交流：Carol Archer, Vicky Cassel, Anne Coyle, Michael Duncan, Jane Forman, Wendy Greene, Erik Jendresen, Gary Kass, Shiva Kumar, Robert Maw-hinney, Pam Morehouse, Pierre Ramond, Amanda Salles和Eero Si-moncelli。我感谢Costas Efthimiou帮我核对事实，寻找资料，还把我的说明正文的草图绘成线条图，而Tom Rockwell又凭他神圣的艺术家的眼光将那些图重新创作出来。我也感谢Andrew Hanson和Jim Sethna，他们曾帮助我准备几幅特殊的图件。

感谢Howard Georgi, Sheldon Glashow, Michael Green, John

Schwarz, John Wheeler, Edward Witten，当然还有Andrew Strominger, Cumrun Vafa和Gabriele Veneziano，他们同意我借用和比较他们各人对不同问题的观点。

我也乐意感谢W.W.Norton出版公司的两位编辑，感谢An-gela Von der Lippe透彻的眼光和珍贵的建议，感谢Traci Nagle对细节的敏锐和认真的态度。我还要感谢我的著作代理人John Brockman和Katinka Matson，他们一直以专家的眼光指引着本书从开篇走到出版。

我做理论物理研究快20年了，要感谢国家科学基金委员会（NSF）、Alfred P.Sloan基金会和美国能源部的大力支持。我自己的研究主要是超弦理论对我们时空概念的影响，这大概也没有什么可奇怪的；在后面的几章，我谈了一些我有幸参与的发现。尽管我希望读者能够喜欢读这些“内线”材料，但我也意识到他们可能会过高估计我在超弦理论发展中所扮演的角色。所以我借这个机会来感谢全世界成千的物理学家，感谢他们为追寻宇宙终极理论所做的贡献。由于所选主题的原因，也因为篇幅的限制，还有很多人的工作没能在书里提到，我向所有那些作者说声抱歉。

最后，我衷心感谢Ellen Archer坚定不移的爱与支持，没有她，这本书是不可能写出来的。

B·格林

# 为B·格林《宇宙的琴弦》喝彩

令人耳目一新的发现源源不断……在物理学家为大众写作的伟大传统中，《宇宙的琴弦》树起一面不倒的旗帜。

George Johnson，纽约时报书评

《宇宙的琴弦》不可不读……霍金为黑洞所做的事情，格林在弦上都做了。

纽约

他写得那么清晰，那么有活力；他以自己的天才，把抽象的物理学原理写活了，也常常把人写乐了。他用别人渴望已久的激情，描绘了超弦理论的纯真，因为那美丽令人动心。

芝加哥论坛报

一本思想性很强的书……《宇宙的琴弦》清晰而迷人地展示了弦理论。它既是个人的故事，也是一场伟大理性运动的故事。

科学美国人

弦理论是自S·霍金关注黑洞以来出现的最有激情的思想……格林用人人都懂的语言解释了弦。

旧金山纪事

格林做了件了不起的事情，用生活的语言解释了弦理论的思想。

它明白如话地说明了那个理论对时空结构的非凡洞察。

新科学家

一本出类拔萃的书。格林为我们多彩的生活带来了一片引人入胜的天地。

自然

B·格林的杰作，是霍金弹起的旋律中的最后（可惜！）一个音符，最美的一个！

伦敦晨星报

非专业的语言，没有一点儿数学，多得惊人的材料……格林清晰简明地写了一部现代科学探险……恐怕没有哪个读者不会被他的激情和兴奋所感动。

费城调查者

太吸引人了……一部辉煌的作品……没有一个方程，人人都能看懂，格林写的谈弦的书，解释了为什么弦会在献身者们中间激起那么大的激情……它让我们能在家中感觉那个抽象得吓人的弦世界，使我们认识到应该认真来看待它。

星期日电讯（伦敦）

格林在用知识、智慧和惊人的鉴别能力写作。

Alan Lightman（《爱因斯坦之梦》作者），哈佛杂志

他的比喻常常使那些原本艰深的概念变得美妙而活泼。《宇宙的琴弦》是一本值得一读的书……爱因斯坦也会满意的。

发现杂志

B·格林让复杂可怕的弦理论走近了每一个人。他凭着惊人的天才，用寻常的语言描绘了可怜的人类感觉以外的维度里可能发生着的事情。

出版者周刊

自《时间简史》的空前成功以来，还没有一本科学读物引起如此的轰动。

星期日时报（伦敦）

B·格林以他动人的文字把外行的人们带到了物理学的前沿。

基督教科学箴言报

一篇来自宇宙学和物理学前沿的（没有方程的）精彩报道。

美国科学家

格林善于阐释最富挑战的科学思想，他的充满洞察的解释会令每一个人耳目一新。

天文学杂志

《宇宙的琴弦》已成为科学解释的经典巨著……弦理论将最终影响我们对美本身的认识。

纽约时报

格林把弦理论带给了广大的读者，揭示了它的含义。那是他穿行在现代物理学的历史和复杂里创造的业绩。

# 第一篇
# 知识的边缘

## 第1章　同一根弦

如果说谁想把事实藏起来，那也太戏剧了。不过，半个多世纪以来，物理学家心里明白，即使对历史上某些最大的科学成就来说，在远方的地平线上也飘浮着乌云。问题在于，现代物理学所依赖的是两大支柱。一个是爱因斯坦的相对论，它为我们从大尺度认识宇宙（如恒星、星系、星系团以及比它们更大的宇宙自身的膨胀）提供了理论框架；另一个是量子力学，我们用这个框架认识了小尺度下的宇宙：分子、原子以及比原子更小的粒子，如电子和夸克。几十年来，两个理论的差不多所有预言都在实验上被物理学家以难以想象的精度证实了。但同样的这两个理论工具，却无情地把我们引向一个痛苦的结论：从广义相对论和量子力学今天的形式看，它们不可能都是正确的。在过去的百年里，我们获得了巨大的进步——解释了宇宙的膨胀，也认识了物质的基本结构——然而，作为这些进步的基础的两个理论，却是水火不相容的。

如果你以前没有听说过这一场火暴的对抗，你也许很想知道那是为什么。这个问题回答起来并不是很困难。除了某些最极端的情形，物理学家研究的东西，要么是小而轻的（如原子和它的组成部分），要么是大而重的（如恒星和星系），从来没有兼具两样性质的。也就是说，对某一样事物，他们只需要量子力学或广义相对论就够了，至于另一家理论怎么大声告诫，可以不屑一顾。50年来，这方法虽然并不令人高枕无忧，但却是非常严密的。

宇宙就可能是极端情形。在黑洞的中央，大量物质被挤压到一个极小的空间里；在大爆炸的时刻，整个宇宙从比沙粒还小的微尘中爆发出来。这些就是“小而重”的领域，体积很小，而质量大得吓人，所以量子力学和广义相对论应该一起走进来。以后我们会越来越明白，当广义相对论与量子力学的方程结合时，会像一辆破车，摇晃、颠簸、丁当哐啷，喷出一路的废气。说白了，那就是，一个良好的物理学问题从两家理论不幸的结合中得到了无聊的结果。即使喜欢让黑洞的内部和宇宙的开端继续躲在神秘背后的人，也不禁会感觉到，量子力学和广义相对论之间的水与火的对抗，只有在更深的层次上才会平息下来。话又说回来，宇宙在最基本的水平上就不能是分离的吗？它也许当真需要拿一组定律来写大东西，而拿另一组不相容的定律去写小的呢。

超弦理论响亮地告诉我们，不是那样的。与令人仰止的量子力学和广义相对论巨人相比，超弦理论不过是初生的牛犊。全世界物理学家和数学家过去十年的研究发现，这种在最基本层次上描写事物的方法，缓解了广义相对论与量子力学间的紧张关系。实际上，超弦带来了更多的东西：在这个新框架下，广义相对论和量子力学的相互需要才使理论有意义。根据超弦理论，“大”定律与“小”定律的结合，不但是幸福的，也是注定了的。

这当然是好事情。而超弦理论——简单说，即弦理论——则将这结合大大往前推进了一步。为了一个能把所有的自然力、所有的物质编织成一幅锦绣图画的统一的物理学理论，爱因斯坦曾追寻了30年。他失败了。今天，在新千年的黎明，弦理论的拥戴者们宣称，那幅迷人的统一图景终于出现了。弦理论有能力证明，发生在宇宙间的一切奇妙的事情——从亚原子世界里夸克疯狂的舞蹈，到太空中飞旋双星高雅的华尔兹；从大爆炸的原初火球，到星河的壮丽旋涡——都体现着一个伟大的物理学原理，一个伟大的数学方程。

弦理论的这些特征要求我们极大地改变对空间、时间和物质的认

识，所以我们需要花一些时间来熟悉它，在一定水平上理解它。不过我们也会发现，从它本来的背景看，弦理论虽然来得突然，却是过去百年物理学革命性发现的自然产物。实际上，我们将看到，像广义相对论和量子力学那样可怕的冲突不是第一次，而是我们在过去百年里遭遇的两次大冲突的结果，那每一次冲突的解决，都使我们对宇宙的认识发生了奇妙的改变。

三次冲突

第一次冲突早在19世纪末就出现了，与光运动的奇特性质有关。简单地说，根据牛顿的运动定律，谁如果跑得足够快，就能赶上远去的光束；而根据麦克斯韦的电磁学定律，谁也跑不过光。我们在第2章会讨论，爱因斯坦通过他的狭义相对论解决了这个矛盾，并因此彻底推翻了我们对空间和时间的认识。根据狭义相对论，空间与时间不再是牢固不变的普适概念，任何人都一样去经历；相反，它们在爱因斯坦的新理论中是以灵活多变的结构出现的，形式和表现依赖于运动的状态。

狭义相对论的发展很快引来了第二次冲突。爱因斯坦理论有个结论说，任何物体——实际上包括任何形式的影响和干扰——都不可能跑得比光还快。但是，正如我们在第3章要讨论的，牛顿那成功经历了无数实验而且大家都感觉满意的引力理论，却牵涉到瞬时通过巨大空间距离的作用。这一次，又是爱因斯坦走上前来，凭他1915年广义相对论的引力新概念，化解了这个矛盾。空间和时间不仅受运动状态的影响，在物质和能量出现时，还会发生弯曲。我们将看到，空间和时间结构的这种扭曲将引力作用从一个地方传到另一个地方。于是，我们不能再把空间和时间看成宇宙万物表现自我的死寂的帷幕；实际上，在狭义相对论和后来的广义相对论中，它们本身也是那些事件的直接表演者。

历史再次重演：广义相对论的发现在解决一个冲突的同时，又带

来另一个。当19世纪的物理学概念用在微观世界的时候，出现了大量令人眼花缭乱的问题，因为这些，自1900年以来的30年里，物理学家们开创了量子力学（在第4章讨论）。前面讲过，那第三个（也是最深刻的一个）冲突，就源自量子力学与广义相对论的水火不容。在第5章我们还将看到，源于广义相对论的弯曲的空间几何形式总是与量子力学蕴含的狂乱的微观宇宙的行为不相容的。到了20世纪80年代中期，弦理论带来一种解决办法，这个冲突才当然地成为现代物理学的中心问题。而且，从狭义和广义相对论成长起来的弦理论，自身也要求严格地修正我们关于时间和空间的概念。例如，我们大多数人都想当然地认为我们的宇宙有3个空间维，但在弦理论看来不是这样的。它认为，宇宙的维数比我们眼睛看到的更多——那些维都紧紧地卷缩在宇宙褶皱的结构中。这个对空间和时间本性的了不起的发现太重要了，在下面，我们将一直用它来做向导。从真正意义说，弦理论讲的就是自爱因斯坦以来的空间和时间。

为理解弦理论到底是什么，我们需要回到从前，简单说说在过去的一个世纪里，我们关于宇宙的微观结构都学会了些什么。

最小的宇宙：关于物质的认识

古希腊人猜想，宇宙的物质是由一些他们叫原子的“不可分割的”原料构成的。他们想，大量的物质都应该是少量不同的基本材料组合的结果，就像在拼音文字里，数不尽的词语都是由那么少的几十个字母组合生成的。这真是先知的猜想。2000多年过去了，我们还认为它是正确的，尽管那些最基本的物质单元已经历了无数认识的转变。19世纪，科学家发现许多熟悉的物质（如氧和碳）都有一种可以识别的最小组成单元，遵照古希腊人的传统，他们称它为原子。名字确定下来了，但历史证明那是一个误会，因为那些原子当然是“可以分割的”。到20世纪30年代初，J·J·汤姆逊（J.J.Thomson）、卢瑟福（Ernest Rutherford）、玻尔（Niels Bobr）和查德威克（James Chadwick）的工作建立了我们熟悉的原子的太阳系模型。原子远不是

什么最基本的物质成分，它有一个包含着质子和中子的核，核外还绕着一群旋转的电子。

有一段时间，许多物理学家都认为质子、中子和电子就是希腊人的“原子”。但是在1968年，斯坦福直线加速器中心的实验家们利用强大的技术力量探索了物质的微观层次，发现质子和中子都不是基本的。反过来，他们证明了那两个“原子”都由3个更小的粒子构成，那些粒子叫夸克——一个古怪的名字，是理论物理学家盖尔曼（Murray Gell-Mann）从乔伊斯（James Joyce）的小说《芬尼根守夜者》里找来的，他早就猜想可能存在着那种粒子。实验家证明，夸克本身有两种，它们的名字不那么有创意，一个叫上，一个叫下。质子由两个上夸克和一个下夸克组成，中子由两个下夸克和一个上夸克组成。

我们在天地间看到的一切事物似乎都是由电子、上夸克和下夸克的组合构成的。没有实验证据说明它们还由更小的东西构成。但却有大量证据表明，宇宙还存在着其他粒子成分。20世纪50年代中期，雷恩（Frederick Reines）和柯万（Clyde Cowan）发现了第四种基本粒子的确凿实验证据，它叫中微子——它的存在，泡利（Wolfgang Pauli）早在20世纪30年代初就预言过了。后来发现，中微子很难找到，它们像幽灵一样，很少与其他物质发生相互作用，能穿透几百亿千米厚的铅，而运动几乎不受影响。这样你可能会感到轻松多了：因为在你读这句话时，太阳向太空喷发的几十亿个中微子正在穿过你的身体，然后穿过地球，继续它们在宇宙间孤独的旅行。20世纪30年代末，物理学家在研究宇宙线（从外太空向地球倾泻的粒子流）时，又发现了一种叫μ子的基本粒子——除了比电子重200倍左右，它们是一样的。μ子在宇宙间的存在，既不是什么事物要求的，也不是人们为解决什么疑问提出的，更不是谁精心设计的。所以，获得诺贝尔奖的粒子物理学家拉比（Isidor Isaac Rabi）对μ子的发现没多大热情，“谁让它来的？”不管怎么说，它来了，而且还跟着来了好多新粒子。

物理学家们凭着前所未有的技术力量，不断地用越来越大的能量

将物质击碎，时刻重现大爆炸以来的那些谁也不曾见过的创生条件。他们从碎片里寻找新的元素，粒子清单越来越长。看看他们发现的东西：另外四种夸克——粲、奇、底和顶——还有一个更重的电子兄弟，τ；另外两个性质与中微子相同的粒子（叫-中微子和τ-中微子，以区别于原来的那个电子中微子）。这些粒子在高能碰撞中产生，不过是昙花一现；在我们通常遇到的任何事物里都没有它们的影子。但是，故事还远没结束。每一个这样的粒子都有一个反粒子——质量相同，而在其他某些方面相反（例如电荷；还有与其他力相应的荷，我们在下面讨论）。举例说，电子的反粒子叫正电子——它的质量跟电子相同，但电荷为+1（而电子的电荷为-1）。物质和反物质接触时，会相互湮灭，生成纯粹的能量——难怪在我们周围的世界里自然出现的反物质会那么少。

物理学家在这些粒子间分辨出一种模式，如表1.1。物质粒子正好分成三组，通常被称为族。每一组包括2个夸克和1个电子，或者电子的伙伴，以及1个相应的中微子。三族中同一行相应的粒子除了质量依次增大而不同外，性质是完全一样的。结果，物理学家现在追溯到了百亿亿分之一米尺度的物质结构，而且证明了我们到目前为止所遇到的每一样事物——不论是自然出现的，还是通过加速器人工产生的——都是由这三族粒子和它们的反物质伙伴组合成的粒子构成的。

| 第 1 族 | | 第 2 族 | | 第 3 族 | |
|---|---|---|---|---|---|
| **粒子** | **质量** | **粒子** | **质量** | **粒子** | **质量** |
| 电子 | 0.00054 | μ子 | 0.11 | τ子 | 1.9 |
| 电子中微子 | $<10^{-8}$ | μ-中微子 | <0.0003 | μ-中微子 | <0.033 |
| 上夸克 | 0.0047 | 粲夸克 | 1.6 | 顶夸克 | 189 |
| 下夸克 | 0.0074 | 奇夸克 | 0.16 | 底夸克 | 5.2 |

表1.1　3族基本粒子及其质量（以质子质量未单位）。中微子质量至今还没有在实验上确定。

从表1.1看，我们对μ子的发现无疑会比拉比更感迷惑。族的划分

至少从表面上显出某种秩序，然而数不清的“为什么”也接踵而来。为什么有那么多基本粒子——特别是，我们周围世界的大多数事物似乎只需要电子、上夸克和下夸克就够了？为什么有三族？为什么不是一组、四族或者更多？为什么粒子质量看起来是随机分布的——例如，为什么$\tau$子比电子重约3520倍？为什么顶夸克比上夸克重40200倍？这些数都很奇怪，似乎是随机数。它们是偶然出现的，还是什么神灵选择的？我们宇宙的这些基本特征能有一个综合的科学解释吗？

力——光子在哪儿

当我们考虑自然力的时候，问题就变得复杂多了。我们的世界充满了施加影响的方式：球拍将球打出，蹦极爱好者从高高的平台跳下，磁体让列车悬浮在金属轨道上飞奔，盖革计数器响应放射性物质时发出“滴答”的声音，原子弹爆炸……我们可以用力推、拉或者摇动物体；可以把一个物体打进另一个物体；可以拉伸、扭转或者粉碎一个物体；还可以令一个物体冷却、加热或者燃烧。在过去的百年里，物理学家积累了大量证据，说明这些不同事物间的相互作用，以及我们寻常遇到的万千事物间的相互作用，都可以归结为四种基本力的组合。其中之一是引力，另外三种力是电磁力、弱力和强力。

引力是大家最熟悉不过的，它不但让我们能牢固地脚踏大地，而且还维持着我们不停地绕着太阳转。物体质量有多大，决定着它能产生多强的引力，对引力会有多大的反应。电磁力在四种力中也是大家熟悉的，它是现代生活中一切方便的动力——例如光、计算机、电视、电话——它在电闪雷鸣时露出狰狞，也在轻轻触摸的手上留下温柔。从微观的角度说，粒子电荷在电磁力中扮演着物质质量在引力中的角色：决定粒子能产生多强的电磁力，对电磁力有多大的反应。

强力与弱力比较陌生，因为它们在超过亚原子尺度以外就完全失去作用了，它们是作用在原子核中的力。难怪这两种力的发现要晚得多。强力将夸克“胶结”在质子和中子内部，又把质子和中子紧紧捆在

一起塞进原子核。弱力最为人所熟悉的作用是物质（如铀和钴）的放射性衰变。

在过去的世纪里，物理学家发现所有这些力有两点共同特征。第一点，我们将在第5章讨论，在微观层次上，所有的力都关联着一个粒子，我们可以把那粒子想象为最小的力元。当我们从“电磁射线枪”射出一束激光时，我们实际是在打开光子的激流，也就是最小的一束电磁力。同样，弱力和强力场的最小单元是一些叫弱规范玻色子和胶子的粒子。（胶子这个名字特别形象，我们可以想象它是把原子核凝结起来的那种强力胶合剂的微观成分。）到1984年时，实验家们已经确定了这三种力的粒子的存在和具体性质，如表1.2。物理学家相信，引力也关联着一种粒子——引力子——不过它的存在还需要实验来证明。

| 力 | 力的粒子 | 质量 |
|---|---|---|
| 强 | 胶子 | 0 |
| 电磁 | 光子 | 0 |
| 弱 | 弱规范玻色子 | 86，97 |
| 引力 | 引力子 | 0 |

表1.2　四种自然力及其相关粒子和质量（以质子质量为单位）。（弱力的粒子有两种可能的质量。理论研究证明引力子是没有质量的。）

第二个共同点是，力由某种“荷”来决定。如质量决定引力如何对粒子产生作用，电荷决定电磁力如何发生影响，粒子还被赋予一定的“强荷”和“弱荷”，它们决定着粒子如何感应强力和弱力的作用（这些性质详细地列在本章注释的表中[1]）。[1]但是，跟质量的情形一样，我们只知道实验物理学家们仔细测量过那些性质，而没有谁能解释为什么我们的宇宙由具有那些特殊质量和力荷的特殊粒子构成。

尽管基本力有这些共同的特征，但是考察它们却只不过使问题更复杂了。例如，为什么有四种基本力？为什么不是五种、三种甚至一种？为什么这些力会有那么多不同的性质？为什么强力和弱力只能在微观尺度上发生作用，而引力和电磁力却具有无限的作用范围？还有，为什么这些力的固有强度会有那么大的悬殊？

为说明最后这个问题，我们想象左手拿一个电子，右手拿一个电子，然后让这两个完全相同的带电粒子靠近。粒子间的相互引力有助于它们靠拢，但电磁斥力却会把它们分开。哪种力更强呢？根本没法儿比：电磁斥力比引力强百亿亿亿亿亿（$10^{42}$）倍！如果说右臂代表引力的大小，那么，为了让左臂能代表电磁力的大小，它必须伸展到我们已知的宇宙边缘的外面。在我们身边，电磁力并没有完全压倒引力，那是因为大多数事物都由等量的正负电荷构成，它们的电磁力相互抵消了。而另一方面，引力总是相互吸引的，不会消减——东西越多，引力就越大。不过根本说来，引力是极端微弱的。（所以，从实验来证实引力子的存在是很困难的。寻找最微弱的力的最小作用单元，是多大的挑战啊！）实验还证明，强力比电磁力强100倍，而比弱力强10万倍。但是，宇宙凭什么有这样的性质——“存在理由”在哪儿？

一件事情为什么恰好是这样而不是那样，是钻牛角尖儿的问题，但我们现在的问题可不是那样的。即使物质和作用粒子的性质稍有改变，宇宙就会大为不同。例如，强力与电磁力的强度比例微妙地决定着构成化学元素周期表上百余种元素的稳定原子核的存在。挤在原子核里的质子因电磁作用而相互排斥，多亏了作用在质子里夸克间的强力才克服了排斥而把它们紧紧系在一起。但是，假如两种力量的相对强度发生极小改变，就很可能破坏它们之间的平衡，使大多数原子核发生分裂。而且，假如电子质量再大几倍，它就会与质子结合成中子，吞噬氢原子核（宇宙间最简单的元素，核里只有一个质子），从而破坏更复杂元素的产生。恒星的存在依赖于稳定核之间的聚变，如

果基本的物理学发生了那些改变，它们也不复存在了。引力的大小也影响恒星的形成。挤压在恒星中心的物质密度是它核熔炉的能源，也是星光的源泉。假如引力的强度增大了，恒星会裹得更紧，从而大大提高核反应的速率。但是，正如烈焰比烛光能更快烧尽燃料，核反应速率的提高会使恒星（如太阳）更快消亡，给我们所知的生命的形成带来致命的灾难。反过来，假如引力强度大大减弱，物质根本就不能聚集在一起，当然更不可能形成恒星和星系了。

我们还可以继续说下去，不过意思已经清楚了：宇宙之所以如此，是因为物质和作用粒子具有那样的性质。但是，它们为什么有那些性质呢？有科学的解释吗？

弦理论：基本思想

弦理论带来了强有力的概念和范式，第一次提出了回答那些问题的框架。我们先来看看基本思想。

表1.1里的粒子是一切物质的基本单元，像语言学里的“字母”一样，它们看来也不会有什么更深层的结构。弦理论却另有说法。根据弦理论，假如我们以更高的精度——比现有技术高许多数量级的精度——去考察那些粒子，我们会发现它们并不是点状的粒子，而是由一维的小环构成的。每一个粒子都像一根无限纤细的橡皮筋，一根振荡、跳动的丝线。物理学家没有盖尔曼那样的文学天才，[2]

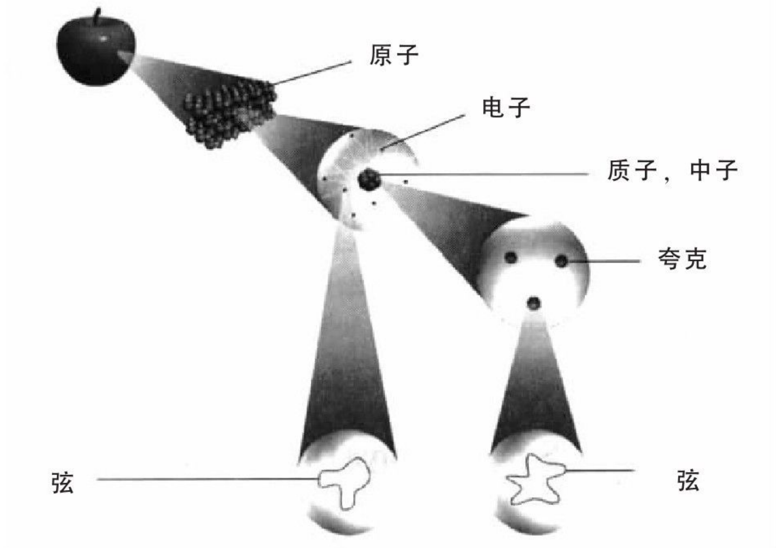



图1.1　物质由原子组成，而原子由夸克和电子组成。根据弦理论，所有这些粒子实际上是振动着的一根闭合的弦。

就把它叫弦。我们在图1.1里说明了弦理论的基本思想：从一个普通的苹果开始，不断地放大它的结构，显出越来越小的组成。以前我们从原子走到质子、中子、电子和夸克，现在弦理论在它前面增添了一根微观的振动的线圈。[2]

虽然不是显而易见的，我们在第6章还是可以看到，单纯地以弦来代替点粒子的物质组成，解决了量子力学与广义相对论之间的矛盾。于是，弦理论解开了当代理论物理学的戈迪乌斯结。[3]这是巨大的成就，但弦理论如此激动人心并不仅仅是因为这一点。

弦理论：万物的统一理论

在爱因斯坦的年代，强力和弱力还没有发现，但他还是为存在两种截然不同的力——引力和电磁力——感到困惑。爱因斯坦不相信大自然会建立在那么奢华的设计图上。于是，他走上了30年的探寻历程，他希望能找到一个统一场论来说明这两种力不过是同一个大的基

本原理的不同表现。这场堂吉诃德式的追求将爱因斯坦从物理学主流里孤立出来。可以理解，物理学家们那时正激动地投入正在兴起的量子力学。爱因斯坦在40年代初给朋友的信里说，“我成了孤独的老头儿，大概主要是因为不穿袜子而出了名，有时候还被当成珍稀动物在特殊场合展览。”[4]

爱因斯坦走在了时代的前头。半个多世纪以后，他的统一理论之梦成为现代物理学的圣杯。很大一部分物理学家和数学家越来越相信，弦理论可能会带来答案。从一个原理出发——万物在最微观的层次上是由振动丝弦的组合构成的——弦理论提供了一个能囊括一切力和物质的解释框架。

例如，弦理论认为，我们观测到的粒子性质，表1.1和表1.2所列的那些数据，不过是弦的不同振动方式的反映。我们知道琴弦（提琴或钢琴）都有共振频率，即弦倾向的振动频率，也是我们耳朵听到的不同的音调与和声——同样，弦理论里的环也有这样的性质。不过，我们将看到，弦理论的弦在共振频率处的振动不是产生什么音乐，而是出现一个粒子，粒子的质量和力荷由弦的振荡模式决定。电子是以某种方式振动的弦，上夸克是以另一种方式振动的弦，等等。在弦理论中，粒子的性质绝非一堆混乱的实验结果，而是同一物理特性的具体表现：基本闭合弦的共振模式——也可以说是弦的音乐。这种思想也适用于自然力。我们将看到，作用力的粒子也关联着特定的弦振动模式，从而天地万物，一切的物质和所有的力都统一到了微观弦振荡的大旗下——那就是弦奏响的“音乐”。

这样，我们在物理学史上第一次有了一个能解释宇宙赖以构成的所有基本特征的框架，因此有时人们说弦理论可能是一个“包罗万象的理论”（theory of everything, T.O.E.）或者是一个“终极”理论。[5]这些浮华的字眼儿不过用来强调那可能是一个最深层的理论——是其他一切理论的基础，而不需要甚至不允许有更基本的理论来解释它。不过，许多弦理论家还是以更老实的态度来看待“万象的理论”，在有限

的意义上思考这个理论有多大能力来解释基本粒子的性质和粒子间相互作用的力的性质。固执的还原论者却认为那不是什么极限，从原则上讲，从宇宙大爆炸到人类幻想的一切事物，都可以用关于物质基本结构的微观物理学过程来描述。在还原论者看来，认识了事物的组成，也就认识了事物本身。

还原论者的哲学很容易激起争论。他们认为，生命和宇宙奇迹不过是循着物理学定律规定的舞步不停舞动着的微观粒子的反映，很多人感到这种观点愚蠢而令人厌倦。难道我们快乐、忧愁和无聊的感觉真的就是发生在大脑里的化学反应吗？——真的是分子和原子间的反应吗？或者，更微观地说，真的是表1.1中的那些原本是振荡的弦的粒子之间的反应吗？为回答这些批评，曾获诺贝尔物理学奖桂冠的S·温伯格（Steven Weinberg）在《终极理论之梦》中告诫说：

反还原论者的另一极端是，他们为其所感觉的现代科学的荒芜感到沮丧。不论他们和他们的世界能在多大程度上还原为粒子的物质或场及其相互作用，他们总觉得被那种认识糟蹋了。……我不想用什么现代科学的美妙来回应那些批评。还原论者的世界观的确是冷漠的，没有一点儿人情味，但我们必须忠实地接受它，不是因为我们喜欢，而是因为世界本来就是那样运行的。[6]

这种鲜明的观点，有人赞同，也有人反对。

还有些人曾试图说明，诸如混沌理论的发展告诉我们，当系统复杂性增大时，会出现一些新的定律来发生作用。认识一个电子或夸克的行为是一回事，用这些知识去理解龙卷风的行为是另一回事。关于这一点，多数人都是赞同的。但是，问题的分歧在于，经常出现在比个别粒子更复杂的系统中的五花八门的意外现象，是否真的说明新物理学原理在发生作用，那些原理是否能够（哪怕是以非常复杂的方式）从统治大量粒子的物理学原理推导出来？尽管很难用电子和夸克的物理学来解释飓风的性质，但我以为那只是计算的尴尬，而不是需

要新物理定律的信号。当然，这一点也有人不同意。

然而，即使我们接受这种有争议的固执的还原论观点，对我们这本书要讲述的历程来说，无疑还存在着严重的问题：原理是一回事，实际是另一回事。几乎所有的人都同意，寻求“一个包罗万象的理论”并不是说要把心理学、生物学、地质学、化学，哪怕物理学的问题都囊括进来解决。宇宙如此丰富多彩，变化万千，我们所谓的终极理论，绝不是科学的终结。恰恰相反，发现T.O.E.——在最微观水平上解释宇宙，而不需要任何更深层的理论来解释它自己——将为我们建立宇宙的新认识提供最坚实的基础。那发现将标志着一个开始，而不是结束。终极理论将为我们树立一座不朽的和谐的纪念碑，它让人们相信，宇宙是可以理解的。

弦理论现状

本书的中心是根据弦理论解释宇宙的行为，特别还要强调这些结果对我们认识空间和时间有什么意义。与其他科学发展的报道不同的是，我们这里讲的理论还没有完成，没有经过严格的实验验证，也没有完全被科学界接受。这是因为弦理论太深奥、结构太精妙，尽管在过去的20年里取得了令人难忘的进步，但离我们完整把握它还着实太远。

所以，弦理论应该看做发展中的理论，而它的部分结果已经带来了令人惊奇的关于空间、时间和物质的新认识。将广义相对论与量子力学和谐地统一起来，是它的主要成功。而且，与其他理论不同，弦理论有能力回答有关自然最基本的物质构成和力的原初问题。同样重要的还有（尽管不太好说），不论弦理论所能提供的答案，还是这些答案的理论框架，都有特别精美的结构。例如，大自然似乎随意表现的那些细节——如不同基本粒子的数目和各自的性质——在弦理论中都是宇宙几何的某些基本而实在的表现。如果弦理论是正确的，我们宇宙的微观结构将是一座错综复杂的多维迷宫，宇宙的弦在其中不停

歇地卷曲、振动，和谐地奏响宇宙的旋律。大自然基本组成的性质绝不是偶然的，而是深刻地与时空结构交织在一起的。

然而，说到底，还得靠确定的可以检验的预言来决定弦理论是否真正揭开了宇宙最深层真理的神秘面纱。要达到那一步，大概还要等一些时候，尽管正如我们将在第9章讨论的，实验验证在未来10年左右能为弦理论提供有力的旁证。而且，我们在第13章会看到，弦理论最近已经解决了一个与所谓贝肯斯坦-霍金熵相联系的有关黑洞的重大难题。20多年来，许多传统的方法都没能解决这个问题。这一成功使许多人相信弦理论正在给我们带来对宇宙行为的最深刻认识。

E·惠藤（Edward Witten）是弦理论的先驱者和卓越的专家，他曾这样概括弦理论的现状："弦理论是21世纪物理学偶然落到20世纪的一个部分"，这话最早是著名意大利物理学家D·阿玛提（Daniele Amati）说的。[7]这样说来，在某种意义上，它有点儿像把一台现代的超级计算机摆在19世纪末的前辈面前，却没有操作指令。通过创造性的反复试验也能显现这台计算机的威力，但要真正把握它还需要更艰辛和长久的努力。计算机的潜在威力跟我们看到的弦理论的强大解释能力一样，将激发人们完全把握它们的强烈愿望。同样的动机在今天正激励着一代理论物理学家去追寻一个精确的解析的弦理论。

惠藤和弦领域的其他专家的言论说明，还要经过几十年甚至几百年我们才可能完全建立和理解弦理论。这很可能是对的。实际上，弦理论的数学很复杂，我们至今也不知道理论的方程是什么。而物理学家只知道那些方程的近似，即使这些近似的方程也够复杂了，只得到部分的解。不过，在20世纪的最后几年出现了一系列激动人心的突破——它回答了迄今难以想象的理论难题——大概预示着我们离完全定量认识弦理论比原先想的要近得多。全世界的物理学家们还在发展比现行各种近似方法更优越的技术，以令人惊喜的速度把弦理论疑惑的分离的元素组织起来。

令人惊奇的是，弦理论的这些发展让我们能够用更好的观点来重新解释一些早已深入人心的理论的基本概念。例如，当我们看表1.1时，会自然生出疑问：为什么是弦呢？为什么不是小飞盘呢？为什么不是一滴滴的小东西？为什么不是这些可能事物的组合？在第12章我们会看到，最近的研究表明，那些事物在弦理论中的确扮演着重要角色，而且，弦理论不过是更宏大的综合理论的一部分——那个理论现在（颇为神秘地）叫M-理论。这些最新发展是我们这本书最后几章的主题。

科学的历程起伏跌宕，有时硕果累累，有时田园荒芜。科学家推出的结果，不论理论的，还是实验的，都摆在科学界同仁的面前，任他们评说。这些结果，有时被否定，有时被修正，有时则为我们重新更精确地认识物理学的宇宙带来思想的飞跃。换句话说，科学曲曲折折地走向我们希望的最后真理，这条路从人类最原始的探索开始，通向我们未知的宇宙尽头。弦理论是这条路上的一个驿站，一个转折点，还是最后的终点，我们不知道。不过，数以百计的来自不同国度的物理学家和数学家们最近20年的研究，使我们能满怀信心地希望，我们正走在正确的道路上，也许离终点不远了。

凭我们现在的认识水平，也能从弦理论获得对宇宙行为的新认识，这一点足以证明弦理论是多么丰富而深刻。我们下面要讲的主要内容是这些理论发展，如何将爱因斯坦狭义和广义相对论开创的空间和时间认识的革命，继续推向前进。我们将看到，假如弦理论是正确的，那么我们宇宙结构的某些性质，也可能令爱因斯坦惊讶万分。

注释

1.下面的表是表1.1的补充。它记录了三族粒子的质量和作用荷。每一类夸克都能携三种可能的强作用荷，我们想象那是夸克的“色”——它代表荷的数值大小。这儿列举的弱荷准确地应该叫弱同位旋的“第3分量”。（我们没有列举粒子的“右手”分量——可以通过

没有弱荷来区别它们。）

2.除了图1.1画的圈（闭弦）外，弦也可以是两端自由活动的（即所谓的开弦）。为表达简洁，我们多数时候都只谈闭弦，不过几乎所有论述都适合于这两种情况。

**族　1**

| 粒子 | 质量 | 电荷 | 弱荷 | 强荷 |
|---|---|---|---|---|
| 电子（e） | 0.00054 | −1 | −1/2 | 0 |
| 电子中微子（$\nu_e$） | $<10^{-8}$ | 0 | 1/2 | 0 |
| 上夸克（u） | 0.0047 | 2/3 | 1/2 | 红，绿，蓝 |
| 下夸克（d） | 0.0074 | −1/3 | −1/2 | 红，绿，蓝 |

**族　2**

| 粒子 | 质量 | 电荷 | 弱荷 | 强荷 |
|---|---|---|---|---|
| $\mu$子 | 0.11 | −1 | −1/2 | 0 |
| $\mu$子中微子（$\nu_\mu$） | <0.0003 | 0 | 1/2 | 0 |
| 粲夸克（c） | 1.6 | 2/3 | 1/2 | 红，绿，蓝 |
| 奇异夸克（s） | 0.16 | −1/3 | −1/2 | 红，绿，蓝 |

**族　3**

| 粒子 | 质量 | 电荷 | 弱荷 | 强荷 |
|---|---|---|---|---|
| $\tau$子 | 1.9 | −1 | −1/2 | 0 |
| $\tau$子中微子（$\nu_\tau$） | <0.033 | 0 | 1/2 | 0 |
| 顶夸克（t） | 189 | 2/3 | 1/2 | 红，绿，蓝 |
| 底夸克（b） | 5.2 | −1/3 | −1/2 | 红，绿，蓝 |

# 第二篇
# 空间、时间和量子的困境

## 第2章　空间、时间和观众的眼睛

1905年6月，26岁的阿耳伯特·爱因斯坦向德国《物理学纪事》投去一篇论文，解决了在少年时代就令他困惑的一个关于光的疑问。杂志的编辑普朗克（Max Planck）在翻过爱因斯坦的最后一页手稿后，意识到大家接受的科学秩序荡然无存了。那位来自瑞士伯尔尼专利局的小职员，已经不声不响地把传统的空间和时间概念彻底推翻了，取而代之的是一个性质与我们在寻常经验中熟悉的任何事物都截然不同的新概念。

困扰爱因斯坦10年的疑惑是这样的：19世纪中叶，苏格兰物理学家麦克斯韦（James Clerk Maxwell）在认真研究了英格兰物理学家法拉第（Michael Faraday）的实验工作后，成功地把电和磁统一在电磁场的框架下。假如你曾在雷雨过后登上山顶，或者站在范德格拉夫发生器的旁边，[8]你对什么是电磁场一定有过切身的体验，因为你已经感觉到它了。假如你还没有那种经历，你可以想象那是电和磁的力线的波浪流过它们所经过的空间区域。例如，当你把铁粉洒在磁铁旁边时，它们形成的有序排列就显示了一些看不见的磁力线。当你在特别干燥的日子脱下羊毛衫时，你会听到“嘶嘶”的声响，可能还会感觉有点儿哆嗦，其实，那就是从羊毛衫纤维脱落下的电荷产生的电力线。麦克斯韦理论不但把这样那样的电和磁的现象统一在一个数学框架，而且还出人意料地发现电磁扰动以恒定不变的速度传播——后来发现，那个速度就是光速。根据这一点，麦克斯韦意识到，可见光不

过是一类特殊的电磁波，我们现在知道，它与视网膜的化学物质发生反应，就产生视觉。另外（这一点很重要），麦克斯韦理论还说明，所有的电磁波都是典型的逍遥客，它们永不停歇，也永不减缓脚步。光总是以光速运动的。

这时还没有什么问题，但问题跟着就来了，那也是16岁的爱因斯坦问过的：假如我们以光的速度追光，会发生什么事情呢？直觉告诉我们，根据牛顿的运动定律，我们将赶上光波，于是光波就像静止不动的——光停在那儿了。然而，根据麦克斯韦的理论和所有可靠的观测，根本没有那样的静止的光；谁也不曾抓一把光在手上。这就是个问题。幸好，爱因斯坦不知道全世界有许多杰出的物理学家正在同这个问题斗争（而且走过许多令人迷失的路线），他在凭着自己独特自由的思路考虑麦克斯韦与牛顿的疑惑。

在这一章里，我们来讨论爱因斯坦如何通过他的狭义相对论解决这个矛盾，如何永远地改变了我们关于空间和时间的概念。也许有人奇怪，狭义相对论首先关心的是，相对运动着的个人（通常叫“观察者”）所看到的世界是什么样的。乍看起来，这不过是没有一点儿意思的智力游戏。事实正好相反，在爱因斯坦的手下，追光的想象隐藏着更深刻的意义。他发现，即使最寻常的事物，在相对运动的观察者看来也会表现最奇异的现象。

直觉和错觉

寻常的一些经验能告诉我们各人看到的事情怎么会不同。例如，路边的树木在驾驶者看来是运动的，而从坐在护栏里等车的人看却是静止的。同样，汽车上的仪表盘在司机看来是不动的（当然不动啦！），但在等车人看来，却是跟着汽车的其他都分一起走的。这些现象太普通，太直观，我们几乎不怎么留意。

然而，狭义相对论认为，不同观察者所看到的现象的不同有着微

妙而深刻的意义。它令人惊奇地指出，相对运动的观察者将感觉不同的距离和时间。我们会看到，这就是说，戴在两个相对运动着的人手上的相同的手表会有不同的节律，从而对任意两个事件之间的时间间隔，也有不同的结果。狭义相对论指出，这个结论并不是说表的精度有问题，它说的是时间本身。

同样，拿着相同皮尺的两个相对运动的观察者将量出不同的距离。这当然还是与他们的测量方法的误差和测量设备的精度无关。世界上最精确的测量仪器也证明，每个人所经历的空间距离和时间间隔是不同的。爱因斯坦的狭义相对论以准确的方式解决了我们关于运动的直觉和光的性质的矛盾，但是也付出了代价：相对运动的观察者不再会看到相同的空间和时间。

自爱因斯坦向世界宣布他那惊人的发现以来，近百年过去了，而我们今天大多数人还在把空间和时间当成绝对的东西。狭义相对论没有深入人心——我们感觉不到它。它的意义在我们的直觉以外。原因很简单：狭义相对论效应依赖于我们的运动速度，而在汽车、飞机甚至宇宙飞船的速度，这些效应是微不足道的。站在地上的人和坐在汽车或飞机上的人的确经历着不同的空间和时间，不过那差别太小而没人注意。然而，假如有人能坐上未来的宇宙飞船以接近光的速度去旅行，相对论效应将变得十分显著。当然，这在今天还是科幻小说的话题。不过，在后面的章节我们将讨论，聪明的实验家们会让我们清楚而准确地看到爱因斯坦理论预言的空间和时间特性。

为实在地感觉上面提到的那些测量，让我们回到1970年，那时刚出现高速的大汽车。斯里姆刚用所有积蓄买了辆新Trans Am赛车，这会儿同兄弟吉姆一道来参加当地的汽车短程加速比赛，想试试那车怎么样（而车商是不会让他们那么试车的）。斯里姆加大油门，汽车飞也似地以每小时120千米的速度跑在那1千米长的跑道上，而吉姆则站在道旁为他测时间。为相互验证，斯里姆自己也拿秒表测量他的新车跑过这段路需要多长时间。在爱因斯坦以前，不会有人怀疑斯里姆和

吉姆会测得完全相同的时间，只要他们的表运行正常。但是依照狭义相对论，如果吉姆的表测得的时间是30秒，那么斯里姆记录的时间将是29.99999999999952秒——小一丁点儿。当然，只有当我们的测量精度远远超过秒表、超过奥运会的计时系统，甚至超过最精确的原子钟，才可能确定那么微小的差别。难怪我们在日常生活中感觉不到时间的流逝依赖于我们运动的状态。

对长度的测量，兄弟俩也会有不同的意见。例如，在下一轮试车时，吉姆用了一种很巧妙的办法来测量斯里姆的新车的长度：当车头经过身边时，打开秒表，车尾经过时，把它按下。因为吉姆知道哥哥的汽车在以每小时120千米的速度前进，所以拿速度乘以他秒表上的时间，就能得到车的长度。当然，在爱因斯坦之前，也不会有人怀疑吉姆以这种直接方法测得的长度与斯里姆在汽车停在车棚里测量的长度是完全一样的。但是，狭义相对论指出，如果兄弟两人用这种办法精确测量了汽车的长度，比如说，斯里姆测得的正好是5米，那么吉姆将发现它是4.99999999999974米——短了一点儿。与时间测量一样，这么小的差别是寻常仪器无法测量的。

差别尽管很小，还是暴露了大众拥有的普适不变的空间和时间概念的致命缺陷。当斯里姆和吉姆的相对速度越来越大时，这缺陷也越来越明显。不过，只有当速度接近最大可能速度（光速）——麦克斯韦理论和实验证明为每秒300000千米——才可能出现可以觉察的差别。那速度足以在1秒钟里绕地球7圈半。如果斯里姆的速度不是120千米/小时，而是9.4亿千米/小时（光速的87%），狭义相对论预言，吉姆测得的车长将是2.4米左右，大大不同于斯里姆的测量（也就是用户手册上标明的长度）。同样，在吉姆看来，短程赛车的时间将比斯里姆测量的时间大1倍。

今天几乎没有东西能达到那样的速度，所以这些专业上所说的“时间膨胀”和“洛伦兹收缩”现象，在日常生活里没有产生什么效应。假如在我们生活的世界里，事物都普遍以接近光的速度运动，那

么空间和时间的这些性质也就完全成了我们的直觉——因为随时都在经历着它——从而就像开头说的路旁的树木那样，也用不着多加讨论了。但是，我们并不生活在那样的世界，所以那些性质还是陌生的。我们会认识到，只有彻底改变自己的世界观，才能理解和接受那些性质。

相对性原理

构成狭义相对论基础的是两个简单然而扎实的原理。一个我们已经提过了，与光的性质有关，在下一节我们还要详细讨论；另一个更抽象，它讲的不是任何具体的物理学定律，却与所有物理学定律都有关系，那就是著名的相对性原理。这个原理基于一个简单的事实：不论我们讨论速度的大小还是方向，都必须明确是谁或者什么在测量。从下面的例子可以很容易地理解这句话的意思和重要性。

让我们想象，在远离星系、恒星和行星的地方，乔治穿着闪红光的太空服漂浮在黑暗的空无一物的空间。从他的角度说，他完全静止地浮在均匀宇宙的黑暗里。他看见远处闪烁着一点绿光，越来越向他靠近。终于，那光走近了，原来是从另一位太空流浪者格蕾茜的太空服发出的，她正慢慢飘过来。经过他时，她向他挥了挥手，他也向她挥挥手。然后，她又消失在黑暗里。这个故事也完全可以从格蕾茜的立场来讲。开始的时候，格蕾茜独自漂浮在太空无边的黑暗中，她看到远处闪烁的红光在向她走来，后来她看清了，那光是从另一个人（乔治）的太空服上发出的。那人向她靠近，经过时也向她挥了挥手，然后消失在远方。[9]

两个叙述以两种不同但等价的观点讲了同一件事情。每个观察者都觉得自己是静止的，而看见别人在运动。每个人的观点都有道理，也都能理解。因为两个太空流浪者的地位是对称的，所以，从根本上说，我们不能讲谁的感觉是“对”还是“错”，他们都一样有理由说自己是对的。

这个例子抓住了相对性原理的精神：运动的概念是相对的。只有在相对于其他事物或与其他事物比较时，我们才能谈一个物体的运动。这样，说“乔治在以每小时10千米的速度运动”是没有意义的，因为我们没有说明任何参照的对象。如果我们以格蕾茜为参照，那么这样讲就是有意义的：“乔治以每小时10千米的速度经过格蕾茜。”正如我们的例子那样，最后这句话完全可以这样说：“格蕾茜以每小时10千米的速度（从相对的方向）经过乔治。”换句话讲，没有“绝对的”运动概念，运动是相对的。

上面的例子中关键的一点在于，乔治和格蕾茜都不以任何方式受力的作用和影响，那些影响可能会改变他们静止的、不受力的作用的、匀速运动的状态。所以，更准确的说法是，只有在与其他对象比较时，不受力的运动才有意义。说明这一点是很重要的，因为，如果出现力的作用，它会改变观察者的速度——改变其大小和（或）方向——而这些改变是可以感觉的。例如，当乔治背着点火的喷气袋时，他准能感觉自己在运动。这种感觉是内在的。只要火箭点火了，乔治就知道他在运动，即使他闭上眼睛，不看周围的事物。即使没有什么比较，他也不会再说自己是静止的，而“其余的世界在他周围运动”。常速的运动是相对的，而非常速的运动（或者说，加速运动）却不是。（我们下一章考虑加速运动，讨论爱因斯坦的广义相对论时，还要回头来检验这种说法。）

让故事发生在太空的黑暗里，更有助于我们的理解，因为它避开了我们熟悉的街道和大厦——我们常常毫无理由地认为它们处在特殊的“静止”状态。实际上，相对性原理也同样适合于地球上的事物，而那也是我们经常遇到的。[1]举例来说，假设你在火车上睡着了，醒来时火车正在通过一段复线。你透过窗户往外看，却被另一列火车挡住了，什么也看不见。这时，你可能说不准是一列火车在动，还是两列都在动。当然，如果火车摇晃或者在弯道上，你会感觉在运动。但如果铁路是笔直的——而且火车速度不变——你只能看到两列火车的相

对运动，而说不准是谁在动。

让我们更进一步。假如你在那列火车上，把窗帘拉下来，把车窗遮住，看不见车厢外的一点儿东西。又假设火车速度是完全不变的，那么，你无法确定自己的运动状态。不论火车停在路上还是高速开着，你看到的车厢都是完全一样的。这样的思想，其实还可以追溯到伽利略，爱因斯坦是通过下面的论断建立起来的：不论是谁，都不可能在这样封闭的车厢里通过实验来决定火车是否在运动。这也是相对性原理：一切不受力的运动都是相对的，只有通过与其他不受力的运动物体或观察者的比较才有意义。如果不与“外面的”事物进行比较，你就不可能知道自己处在什么运动状态。根本没有什么“绝对的”匀速运动，只有比较才有物理意义。

实际上，爱因斯坦发现，相对性原理还有着一个更响亮的论断：不论什么物理学定律，对所有匀速运动的观察者来说都是完全相同的。假如乔治和格蕾茜不是孤零零地漂浮在太空，而是在各自的太空站里做同一组实验，那么他们的实验结果还是相同的。每个人都一样有理由相信自己的太空站是静止的，尽管两个站是相对运动着的。如果他们所有的仪器都一样，两个实验室就没有什么分别——是完全对称的。他们从实验得出的物理学定律也是相同的。无论他们自己还是他们的实验，都不可能感觉到匀速运动——也就是说，不以任何方式依赖于那种运动。就是这个简单的概念在两个观察者之间建立起完全的对称关系；就是这个简单的概念体现了相对性原理的精神。很快我们会将这个原理用于重大的效应。

光速

狭义相对论的第二个关键因素与光和光的运动性质有关。我们说过，“乔治以每小时10千米的速度运动”这句话离开比较对象是没有意义的，然而光却不同。一个世纪以来，大量实验物理学家的努力都证明，一切观察者都同意光以每秒300000千米的速度运动——不论以

什么标准为参考。

这个事实变革了我们的宇宙观。为弄懂它的意义，我们先来看，那个关于光速的论断对普通的事物是不是对的。想象一下，在一个明媚的日子里，你跟朋友出去玩儿沙滩排球。你们快乐地把球传来传去（速度比如说是每秒6米）。忽然，天上电闪雷鸣，你们赶紧跑去找躲雨的地方。雨过天晴，你们又重新玩儿起来。可是你发现有点不对劲儿，朋友的头发乱蓬蓬的，两眼变得凶恶而疯狂。再看她的手，你惊奇地发现她手上拿的不是什么球，而是要把一颗手榴弹扔给你。当然了，你玩球的热情一下子烟消云散了，转身拔腿就跑。你的伙伴扔出手榴弹向你飞过来，但因为你也在跑，所以它向你追来的速度比6米/秒小。实际上，经验告诉我们，如果你跑的速度是每秒4米，那么手榴弹向你飞来的速度是（6-4＝）2米/秒。再看一个例子：假如你在山上忽然遭遇雪崩，你首先想到的是跑，因为那样雪向你压过来的速度会慢下来——这当然是好事。同样，静止的观察者看到的雪速度要比逃跑者感觉的快。

现在，我们来比较一下排球、手榴弹、雪崩与光有哪些基本差别。为了让比较更密切，我们想象光是由一“束束”或一“包包”光子组成的（光的这点性质我们在第4章还要更详细地讨论）。当我们打开手电筒或者激光器时，实际上就在向某个方向发射光子流。像手榴弹和雪崩的例子一样，我们来看，运动的观察者看到的光子是如何运动的。假定你那位发了疯的朋友把手榴弹换成大功率的激光向你射过来——你可以发现（假如你有很好的测量仪器），光子束的速度为每小时10.8亿千米。但是，假如你像看到手榴弹飞过来时拔腿就跑，情况会怎样呢？光向你飞来的速度会是多大呢？为了更令人相信，请你坐上“冒险者”号飞船，以1.6亿千米/小时的速度逃离你的伙伴。这样，照传统的牛顿世界观，你大概以为光子飞向你的速度会慢一些，因为你也在跑。具体地说，你预料它们向你靠近的速度是（10.8-1.6＝）9.2亿千米/小时。

自1880年以来，大量不同的实验以及对光的麦克斯韦电磁学理论的分析和解释，逐渐令科学家们相信，你不会看到你想象的那种事情。实际上，不论你怎么跑，你总会发现光子向你飞来的速度是10.8亿千米/小时，一点儿也不会慢。乍听起来，这似乎很荒唐，一点儿也不像我们在排球、手榴弹和雪崩时发生的事情。然而，事实就是那样。不论你迎着光还是追着光跑，它靠近或离开你的速度是不会改变的，都是每小时10.8亿千米。不论光子源与观测者如何相对运动，光速总是一样的。[2]

由于技术的局限，上面说的那些“实验”实际不可能完成。不过，比较的实验还是可以做的。例如，荷兰物理学家德西特（Willem de Sitter）在1913年提出，快速运动的双星（两颗相互绕对方旋转的恒星）可以用来测量光源的运动对光速的影响。80多年来，许许多多的这类实验都证明来自运动恒星的光与来自静止恒星的光具有相同的速度——在不断提高的仪器精度下，都是10.8亿千米/小时。另外，在过去的百年里，在不同环境下做了许多直接测量光速的实验，还检验了光的这种性质所带来的许多结果——它们都证明，光速是一个常数。

如果你觉得光的这种性质很难理解，那不是你一个人的问题。在19、20世纪之交的那些年，曾有许多物理学家想尽办法来反对它，但都失败了。爱因斯坦不一样，他欣然接受了不变的光速，因为它解决了困惑他10多年的矛盾：不论你怎么费力去追赶，光总是以光速跑在你的前头。你不可能觉察光速有一丁点儿的差别，当然更不可能让光慢慢停下来。问题解决了，但不仅仅是战胜了一个难题。爱因斯坦发现，不变的光速意味着牛顿物理学的崩溃。

事实和结论

速度度量一个物体在一定时间间隔内能走多远。如果坐在速度为65千米/小时的汽车上，我们在1小时里当然走了65千米（只要在这个

小时内我们保持相同的运动状态）。这样说来，速度是很普通的概念。那么有人可能奇怪，我们为何还费大力气去谈什么排球、雪球和光子的速度。但是请注意，距离是关于空间的概念——特别是它度量了两点间有“多少”空间。另外还应注意，间隔是关于时间的概念——两个事物之间经历了多长时间。于是，速度最终是与我们的空间和时间概念联系着的。这样我们看到，挑战我们寻常的速度概念的那些实验事实，如光速的不变性，实际上也在挑战我们寻常的空间和时间概念本身。因为这一点，光速的奇特性质值得更仔细地研究——爱因斯坦通过对它的考察，得到了惊人的结果。

同时性

根据光速的不变性，可以毫不费力地证明我们平常熟悉的时间概念是完全错误的。假定有两个敌对国的元首，分别坐在长长的谈判桌的两头。他们刚达成停战协议，可谁也不愿先在协议上签字。联合国秘书长走过来，他想到一个绝妙的解决办法。把一盏灯放在桌子的中间，灯光会同时到达两位总统（因为他们距离灯是一样远的）。当两个总统看到灯光时，就在协议文本上签字。就这样，协议在双方都满意的情况下达成了。

秘书长很高兴，又用同样的办法来调解另外两个正在备战的国家。不同的是，谈判在匀速行驶着的火车上进行。两个国家的总统坐在谈判桌的两头，“前卫国”总统面对火车前进的方向，“后卫国”总统面对他坐在对面。秘书长知道，只要运动状态保持不变，物理学定律就总是一样的，而与各人的运动状态无关，所以谁坐在哪头是没有关系的。他又主持了那种“灯光签字仪式”。两位总统签署了协议，与幕僚们共同庆祝两国结束敌对关系。

这时候，有人来报告，在车外站台上看签字仪式的两国群众打起来了。谈判列车上的人很震惊，他们听说两国群众冲突的原因是“前卫国”的人感觉自己受骗了，因为是他们的总统先在协定上签了字。

而车上的人——不论哪一方——都认为签字是同时进行的。外面的人怎么会看到不同的场面呢？

让我们更仔细地来考虑站台上的人所看到的情形。当初，谈判桌上的灯是关着的，然后在某个时刻打开，光传向两位总统。从站台看，“前卫国”总统迎着照过来的光，而“后卫国”总统则在离开光。这就是说，对站台上的人而言，灯光离“前卫国”总统的传播路线比离“后卫国”总统的更短，因为一个迎着光来，一个离光而去。这不是说光的速度在射向两位总统时有什么不同——我们已经讲过，不论光源和观察者的运动状态如何，光速都是相同的。我们这里说的只是，从站台上的观察者的观点看，光到达两个总统所经历的距离有多远。因为光到“前卫国”总统的距离比到“后卫国”总统的短，所以它将先达到“前卫国”的总统，这就是为什么“前卫国”的公民说自己上当了。

当有线新闻网（CNN）广播群众看到的情景时，联合国秘书长、两国总统以及幕僚们都惊呆了，简直不敢相信自己的耳朵。他们都看到灯肯定是精确地放在两位总统的正中央的，如果没有什么干扰，灯发出的光传到他们的距离是一样的。因为光向左和向右的速度相同，他们相信——而且确实看到了——光真的是同时到达两个总统的。

车上车下的人，谁对谁错呢？双方看到的和解释的理由都无懈可击。答案是，两家都是对的。像乔治和格蕾茜那两位太空行者的情形一样，两种观察结果都有理由说是正确的。惟一令人疑惑的是，这里的两种情形似乎是相互矛盾的。出现了棘手的政治问题：两位总统是同时签字的吗？以上面的观察和理由使我们不得不相信，根据列车上的人的观点，他们是同时签字的，而根据站台上的人的观点，他们不是同时签字的。换句话讲，如果两个观察者是相对运动的，那么在一个人看来同时发生的事情，在另一个人看来是不同时的。

这是一个惊人的结论，是对实在本性最深刻的洞察之一。不过，即使多年以后你忘了这一章讲的事情，而还能记得那艰难的和平历

程，那么你还是把握了爱因斯坦发现的精髓。时间的这种出人意外的性质，不需要令人皱眉的数学，也不需要眼花缭乱的逻辑，它是光速不变性的直接结果，这一点我们已经说过了。我们现在来看，如果光速不是常数，而像我们直觉认为的那样，像排球、雪球的速度那样变化，那么站台和列车上的人的意见就不会有冲突了。站台上的人还是会说，光离“后卫国”总统的距离要比离“前卫国”总统远一点儿，但直觉告诉我们，光飞向“后卫国”总统的速度也要快一点儿，因为向前奔驰的火车也在给它“加劲儿”。同样，他们会看到光飞向“前卫国”总统的速度会慢一点儿，因为向前的列车会将它“拖住”。考虑了这些效应（当然是错误的），站台上的人们会看到光同时到达两位总统。然而，在现实世界里，光不能被加速，也不会慢下来，所以火车既不可能使它更快，也不可能使它变慢。于是，站台上的观察者最终还是会说光先到达“前卫国”的总统。

千百年来，我们一直以为同时性的概念是普适的，不论运动状态如何，都是大家公认的；然而，光速不变性要求我们放弃这种观念。我们曾经幻想一种普适的时钟，不论在地球、火星、木星还是在仙女座星系，它在宇宙的每一个角落都能以完全相同的节律，一分一秒地走下去。现在看来，这样的钟是不可能存在的。反过来说，相对运动的观察者对事件是否同时发生，会有不同的看法。然而，还是因为我们寻常遇到的速度太小，所以我们世界的这种实实在在的特征对我们来说依然是陌生的。假设谈判桌长30米，火车以每小时16千米的速度运行，那么站台上的人们会“看到”光到达“前卫国”总统的时间比达到“后卫国”总统的时间要早大约一千万亿分之一秒。虽然这是真正的差别，但确实太小了，我们不可能直接感觉得到。假如火车快得多，每小时跑10亿千米，那么，站台上的人会看到光到达“后卫国”总统的时间要比到“前卫国”总统多20倍。在高速情况下，狭义相对论的惊人效应就越发显著了。

## 时间的膨胀

很难为时间下一个抽象的定义——那常常会把“时间”本身卷进来，要不就得在语言上兜圈子。我们不想那么做，而采取一种实用的观点，将时间定义为时钟所测量的东西。当然，这也把定义的负担转给了“时钟”。这里，我们不那么严格地将时钟理解为一种做着完全规则的循环运动的仪器。我们通过计数时钟经过的循环次数来测量时间。像手表那样的寻常钟表是满足这个定义的，它的指针规则地一圈一圈地转，而我们也的确通过它的指针在两个事件之间转的圈数来确定时间。

当然，“完全规则的循环运动”也隐含着时间的概念，因为“规则”指的正是每一个循环经历相同的时间间隔。从实用的立场出发，我们用简单的物理过程来建立时钟，就是说，我们希望它在原则上反复地循环，从一个循环到下一个循环不发生任何方式的改变。古老的来回摇荡的摆钟和以重复的原子过程为基础的原子钟，为我们提供了简单的例子。

我们的目的是认识运动如何影响时间的流逝。既然我们已经以操作的方式用钟的运动定义了时间，那么也可以将问题转换为：运动如何影响钟的“嘀嗒”？首先应强调一点，我们的讨论并不关心某个特殊的钟的机械零件会在摇晃、碰撞中发生什么事情。其实，我们要讲的只是最简单最平凡的运动——速度绝对不变的运动——这样也不会有摇晃或碰撞。我们真正感兴趣的是一个普遍性的问题：运动如何影响时间的流逝，也就是说，如何根本地影响任何钟的节律，而与钟的具体设计和构造无关。

为此，我们引入一种最简单的概念性的（不过也是最不实用的）钟，那就是所谓的“光子钟”。它由安在架子上的两面相对的小镜子组成，一个光子在两面镜子间来回反射（图2.1）。假定镜子相隔15厘米，光子来回一趟需要大约十亿分之一秒。我们可以把光子的一次来回作为光子钟的一声“嘀嗒”——嘀嗒10亿声就意味着经过了1秒。

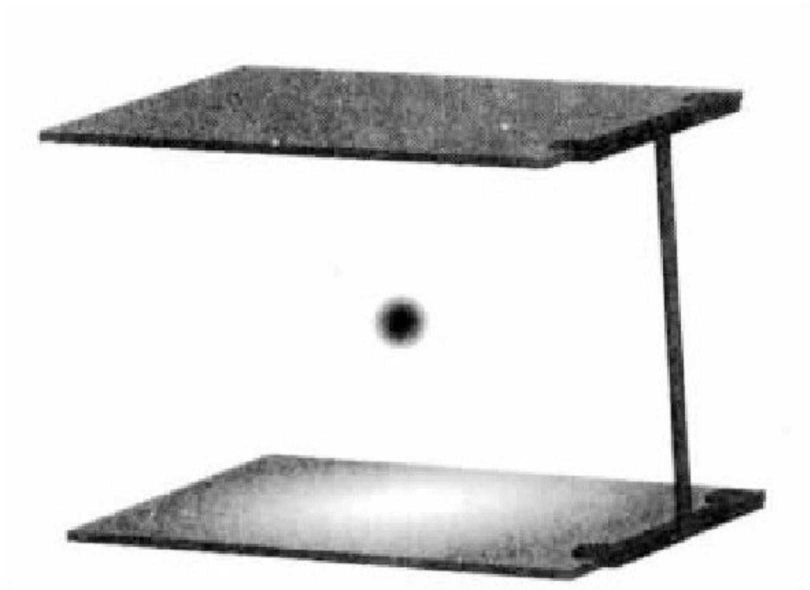


图2.1　两面平行镜子构成的光子钟，中间有一粒光子。光子每完成一次往返，钟就“嘀嗒”一声。

我们可以拿光子钟做秒表来测量两个事件的时间间隔：只需要数一下在我们感兴趣的期间里听到了多少次“嘀嗒”声，然后用它乘以每次“嘀嗒”所对应的时间。例如，我们测量一场赛马的时间，从开始到结束，光子来回的次数为550亿次，那么我们知道赛马经过了55秒。

我们用光子钟来讨论是因为它的力学性质很简单，而且摆脱了许多外来的影响，从而能让我们更好认识运动如何影响时间过程。为看清这一点，我们来仔细看看身边桌上的光子钟是怎么计时的。这时候，忽然从哪儿落下另一只光子钟，在桌面上匀速地滑过（图2.2）。我们的问题是，运动的钟与静止的钟会以相同的节律“嘀嗒”吗？

为回答这个问题，让我们从自己的角度来看光子在滑动的钟内为了一声“嘀嗒”该走的路径。如图2.2，光子从滑动着的钟底出发，然后到达上面的镜子。在我们看来，钟是运动的，光子的路径应该像图

2.3那样是斜的；如果光子不走这条路，就会错过上面的镜子而飞向空中。然而，滑动的钟也有理由说自己是静止的而其他东西在运动，我们也知道光子一定会飞到上面的镜子，所以我们画的路线是对的。然后，光子从上面反射下来，沿着另一条斜线落回下面的镜子，“敲响”滑动的钟。显然，我们看见的光子经历的两条斜线比光子在静止的钟里从上到下的直线更长，因为从我们的视角看，光子不仅上下往返，还必须随滑动的钟从左飞到右。这一点是有根本意义的。另外，光速不变性告诉我们，滑动钟的光子与静止钟的光子一样，都以光速飞行。光子在滑动的钟里需要飞过更长的路径，所以它“敲响”的钟声会比静止的钟少。这个简单的论证说明，从我们的视点看，运动着的光子钟比静止的光子钟“嘀嗒”得慢。而我们已经认为“嘀嗒”的次数反映了经历时间的长短，因此我们看到，运动的钟的时间变慢了。

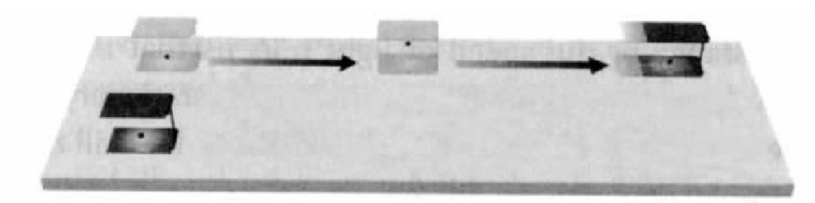


图2.2　前面是静止的光子钟，另一只光子钟匀速滑过。

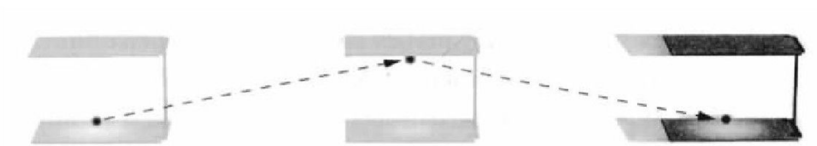


图2.3　从我们的视点看，光子在滑动的钟里走过一条折线。

你可能想问，也许这不过是光子钟的特殊性质，未必适合于古老的摆钟或者劳力士手表。这些更熟悉的钟表测得的时间也会慢吗？我们可以响亮地回答“是的”，可以用相对性原理来证明。在光子钟上系一只劳力士表，重复刚才的实验。我们已经讲过，静止的光子钟和系

在上面的劳力士表所测量的时间是一样的，光子钟“嘀嗒”10亿次，劳力士表走1秒钟。如果光子钟和劳力士表在运动呢？劳力士表会像光子钟那样也同步地慢下来吗？为使问题更明白，我们把钟和表固定在列车车厢的地板上，车厢没有窗户，列车在笔直光滑的铁路上匀速地滑行。根据相对性原理，车上的人谁也没有办法判断列车是否在运动。但如果劳力士表和光子钟不同步，他们就可以凭这一点发现运动的效应。因此，运动的钟和钟上的表一定测量相同的时间间隔；劳力士表一定以完全相同的方式像光子钟那样变慢了。不论什么牌子、什么类型、什么结构的钟表，只要在相对运动，它们就会测量出不同的时间节律。

光子钟的讨论还说明，静止与运动的钟的时间差决定于滑动钟的光子完成一次往返飞行需要经过的距离，而这又决定于钟滑动的速度——从静止的观察者看，钟滑动越快，光子飞行越远。所以，与静止的钟相比，滑动的钟滑得越快，它“嘀嗒”的节律就越慢。[3]

为了对时间大小有一点感觉，我们注意光子来回一趟大约是十亿分之一秒。能在“嘀嗒”声中经过一段可以觉察的路径的钟一定运动得很快——那速度与光速差不多。假如它以寻常的每小时16千米的速度运动，则在光子走完一个来回时它才移动了五百亿分之一米。这个距离太小，从而光子经过的距离也小，相应地，对钟的影响也小了。根据相对性原理，这同样适合于所有的钟——也就是说，适合于时间本身。这也是为什么我们这些以低速度相对运动的生命一般都感觉不到时间的扭曲。那效应虽然肯定存在着，却是小得惊人。相反，假如我们能抓着滑动的钟，跟它一起以3/4光速运动，那么我们可以用狭义相对论方程证明，静止的观察者会发觉我们运动的钟的节律大约只是他们的钟的2/3，这实在是显著的效应。

运动的生命

我们看到了，光速不变性意味着运动的光子钟比静止的光子钟

的“嘀嗒”节律慢，而根据相对性原理，这不仅对光子钟是正确的，也适合于任何类型的钟——也就一定适合于时间本身。运动的观察者的时间过得比静止观察者的慢。照这样的推理，运动着的生命岂不是比静止的生命活得更长吗？毕竟，假如运动的时间比静止的时间慢，那么这种差别不应仅适用于钟表测量的时间，也同样适用于心跳和身体器官衰老所决定的时间。真是这样的，它已经得到了直接证实——不是人的寿命延长了，而是来自微观世界的某种粒子（μ子）的寿命延长了。然而，我们却不能说找到了青春的源泉，因为面前还有巨大的困难。

在实验室里处于静止状态时，μ子经过类似放射性衰变的过程，在大约百万分之二秒的时间内发生分裂，这是得到无数证据证明了的实验事实。μ子仿佛举着一支枪顶着自己的头，在百万分之二秒时扣动扳机，把自己击碎，分裂成电子和中微子。但是，假如μ子不是静止在实验室里，而是在某个粒子加速器里，它将获得只比光慢一点的速度。实验室的科学家会发现它的平均寿命惊人地延长了。确实如此。以每小时10.73亿千米（约99.5%的光速）运动的μ子，寿命大约会增大10倍。照狭义相对论的解释，快速运动的μ子“戴”的表比实验室里的钟慢得多，当实验室钟声响起该它开枪时，它的表还远没到那“最后的时刻”。这说明了运动对时间过程直接而惊人的影响。假如谁能像μ子那么快地飞翔，他的生命也会一样地延长。原先活70岁的，会活到700岁。[4]

现在我们来看那困难是什么。实验室的人看到高速运动的μ子比它静止的伙伴活得更长，那是因为对运动者来说，时间走慢了。不仅μ子的表慢了，它经历的一切活动都慢了。例如，假如静止的μ子一生能读100本书，它那运动的兄弟也只能读100本书——尽管它的寿命长多了，但它阅读的速度和它生命的一切活动也都慢下来了。从实验室看，运动的μ子会比静止的活得更长，但它经历的“生命的总和”却是一样多。这个结论当然也适用于那些高速运动的能活几百岁的人。

在他们自己看来，生命如故。在我们看来，他们过着超慢节奏的生活，他们的一个普通生命周期要经历我们漫长的时间。

谁在运动

运动的相对性既是爱因斯坦理论的钥匙，也是混乱的根源。你大概已经注意了，如果换一个角度看，那么时间过得慢的“运动的”μ子与“静止的”μ子将相互改变角色。像乔治和格蕾茜都能说自己是静止的一样，我们讲的运动的μ子完全可以从它自己的角度说它没有动，真正（在相反方向）动的是那“静止的”伙伴。从这个角度看，前面的论证同样是成立的，于是我们得到一个表面上很矛盾的结果：我们所讲的静止的μ子的时间，相对于我们所讲的运动的μ子的时间，慢了。

在“灯光签字仪式”的例子中，我们曾遭到过这样的情形：不同观点会带来离奇的结果。在那里，我们被迫根据狭义相对论的基本论证放弃了这样一个根深蒂固的旧观念：不论在什么样的运动状态，人们对事件发生的时间会有一致的认识。而眼前的冲突似乎更严重。两个观察者怎么可能都说对方的表慢了呢？更令人惊讶的是，两个不同的μ子的观点使我们面对这样一个严酷而悲哀的境地：两个兄弟都说自己会先离开这个世界。我们知道世界上会发生一些出人意料的怪事，但我们还是不希望出现逻辑荒唐的事情。究竟是怎么一回事呢？

像狭义相对论出现的其他悖论一样，仔细考察这些逻辑怪圈会带来对宇宙行为的新认识。为避免过分的拟人化，我们不谈μ子兄弟了，还是来看乔治和格蕾茜。现在，他们除了太空服上的闪光灯以外，还带着明亮的数字钟。在乔治看来，他是静止的，而格蕾茜的灯光和她巨大的数字钟出现在远处，然后从黑暗的虚空空间走过来，经过他。他发现她的钟比自己的慢（慢多少则依赖于他们相互经过的速度是多大）。如果再机灵一点儿，他还会发现，不仅她身上的钟慢了，她的一切都慢了——她经过时挥手的动作慢了，她眨眼睛的速度慢了……在格蕾茜来看，这些缓慢的运动同样发生在乔治身上。

虽然这显得很奇怪，我们还是来看一个揭示逻辑荒谬的精确实验。最简单的办法是，让乔治和格蕾茜相遇时把他们的钟都拨到12：00。两人分开后，都说对方的钟慢了。为看个究竟，他们只好又回到一起，直接比较钟的时间。不过，他们怎样才能再相遇呢？既然乔治带着喷气袋，他当然可以利用它来追格蕾茜（从他的角度看）。但是，如果他真那么做，那引发悖论的两人的对称关系就被破坏了，因为乔治现在经历着加速，而不是没有力作用的自由运动。当他们这样重逢时，乔治的钟真的慢了，他可以肯定地说自己在运动，因为他感觉到了。乔治与格蕾茜的观点不再相同。打开喷气袋时，乔治就不再说自己是静止的了。

假如乔治就这样追赶格蕾茜，他们的相对速度和乔治喷气的具体方法将决定两人的时间会有多大差别。我们现在已经知道，如果相对速度小，时间差别也会很小；如果速度同光速差不多，则时间差可能会是几分钟、几天、几年、几百年，甚至更大。考虑一个具体的例子：乔治和格蕾茜以99.5%的光速分离，3年以后（据乔治的钟），乔治在瞬间点燃他的喷气袋，以同样的速度去追格蕾茜。当他追上她时，6年过去了——这是他的钟所经历的时间，因为他需要3年才赶得上格蕾茜。然而，狭义相对论的数学证明，格蕾茜的钟这时已过了60年。这不是什么梦幻：格蕾茜得追寻60年前的记忆，才会想起她经过乔治的那一刻。而对乔治来说，那不过是6年前的事情。乔治真的成了时间行者，准确地说，他走进了格蕾茜的未来。

让两个钟回到一起来面对面地比较，这似乎只是一个逻辑小把戏，然而的确触及了问题的核心。我们想过很多办法来克服这点疑惑，最终都失败了。例如，我们不让钟回到一起，而让乔治和格蕾茜通过网络电话联系来比较他们的时间，事情会如何呢？如果这种联系是瞬间的，我们就不得不面对一个难以逾越的障碍：从格蕾茜的角度看，乔治的钟走得较慢，所以他通报的时间一定会小些；从乔治的角度看，格蕾茜的钟走得更慢，所以她通报的时间一定会小些。两个人

不可能都是对的，而我们却糊涂了。问题的关键在于，网络电话同所有通讯方式一样，不是瞬时传递信号的。电话经过无线电波（光的一种）传达信号，因此信号也以光速传播。这意味着接收信号需要一定的时间——实际上，正是这一时间延迟，将彼此的观点协调起来了。

我们先从乔治的角度来看。假定在每小时正点的时候，乔治就在电话里报告，“现在是12点正，一切正常”，“现在是1点正，一切正常”……在他看来，格蕾茜的钟走得慢，所以他开始以为她的钟在她收到通话后还没走到那个钟点。于是，他认为，格蕾茜会同意她的钟走得慢。但他马上又想，“格蕾茜在离我而去，我给她的电话一定要经历更远的距离才能到达她。也许，这多出的传话时间正好补上她走慢的钟。”乔治想到了存在着两种对立的效应——一方面，格蕾茜的钟走得慢；另一方面，他的信号传播需多费些时间——于是，他满怀热情地坐下来计算这两个效应的综合结果。他发现，传播信号需要的时间超过了格蕾茜的钟慢的那段时间。结论令他惊讶：格蕾茜要在她的钟过了那点以后才能收到他报告那点的电话。实际上，乔治知道格蕾茜也精通物理学，知道她在根据他的电话确定他的时间时，会把信号传播的时间考虑进来的。经过更多的计算，我们会证明，即使考虑了信号传播的时间，格蕾茜在分析他的信号后，也会得到这样的结论：乔治的钟比她的慢。

从格蕾茜的角度看，让她向乔治发正点报时信号，上面的论证也一样适用。起初，她觉得乔治的钟走得慢，因而会在他到点以前收到她的正点消息。但她接着考虑了她的信号一定要走得远一些才能追上正消失在黑暗中的乔治，于是她意识到，他实际上会在发出自己的正点信号以后才能收到她的消息。她同样意识到，即使乔治考虑了信号传播的时间，他也会根据她的电话得到结论：她的钟比他自己的慢。

只要乔治和格蕾茜都没有加速，他们两个的观点就都是站得住脚的。尽管表面看来像一个悖论，但他们却通过这种方式发现，在认为对方的钟走得慢这一点上，他们是完全一致的。

空间的收缩

上面的讨论说明观测者会看到运动的钟比自己的钟走得慢——就是说，运动影响时间。向前一小步，我们可以看到运动也同样惊人地影响着空间。我们回头来看斯里姆兄弟和他们的短距离试车。我们讲过，当汽车还停在展厅时，斯里姆就用皮尺认真测量过新车的长度。当汽车在跑道上飞驰时，吉姆不可能再用皮尺去量，只好用一种间接的办法。我们在前面曾提过一个办法：在车头经过时，吉姆打开秒表；车尾经过时，吉姆按下秒表，然后，用这个时间乘以汽车的速度，就能确定车的长度。

根据刚发现的时间特性，我们知道，在斯里姆看来，自己是静止的，吉姆是运动的。于是，他看到吉姆的钟走慢了。结果，斯里姆认为吉姆用间接方法测量的汽车长度比他自己在车库测量的短。因为，在吉姆的计算里（车长等于速度乘以经过时间），他用的是走得慢的表测量的时间。既然表慢了，他看到的时间短了，他计算的结果一定也短了。

于是，吉姆将感觉斯里姆的汽车在运动中会变得比在静止时短。这不过是一个例子。一般情况下，观察者会看到运动的物体在运动方向上缩短了。[10]例如，狭义相对论的方程证明，如果物体以90%的光速运动，那么静止的观察者将发现它比静止时短了80%。图2.4画出了这个现象。[5]

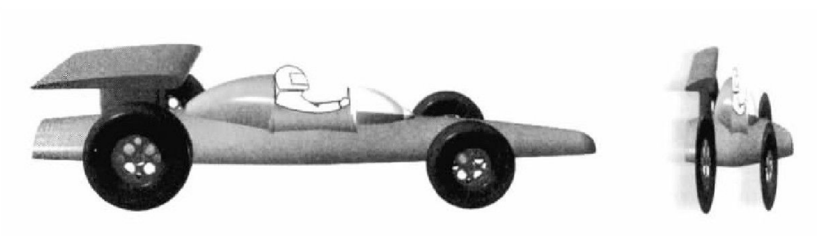


图2.4　运动物体在运动方向上缩短了。

在时空里运动

光速不变性引出一种新的空间和时间概念，取代了传统的固定的刚性结构的空间和时间观念。在新的概念里，空间和时间的结构密切依赖于观察者和被观察者之间的相对运动。我们已经认识到运动物体的演化慢了，在运动方向上的长度也缩短了，本可以就这样结束这儿的讨论。然而，狭义相对论还提供了一个更为深刻的统一的观点，能囊括所有这些现象。

为了理解这个观点，我们想象有辆不那么现实的汽车，能很快达到省油速度，160千米/小时，然后保持这个速度，不快也不慢，最后突然刹车停下来。这时候，斯里姆高超的驾驶技艺越来越出名了，于是人们请他在广袤平坦的大沙漠上的一条笔直的路上试开这辆汽车。从起点到终点，路线长16千米，汽车6分钟（1/10小时）就能开过去。吉姆这回充当汽车技师，检查12组试车数据。令他困惑的是，尽管多数记录的时间都是6分钟，但最后三次却长一些：6.5分、7分、7.5分。起初他怀疑是机械故障，因为这几个时间说明汽车在最后三轮的试验中速度没能达到160千米/小时。但是，认真检查后，他相信汽车没有一点儿问题。他无法解释那些反常的时间，就去问斯里姆最后三轮的情况。斯里姆的解释很简单。他告诉吉姆，在最后三轮，天近黄昏，车从东头开向西头，他的眼睛正对着落山的太阳，于是把车开偏了一点儿。他还画了一张草图说明最后三轮的路线（图2.5）。现在明白了为什么那三轮的时间会长一些：从起点到终点偏了一个角度，路线更长了，因而相同的速度需经历更长的时间才能开过去。换句话说，当路线偏离一个角度时，160千米/小时的速度有一部分耗在了从南到北的方向上，于是从东到西的速度就慢了一点儿，从而经过这段路线的时间会长一点儿。

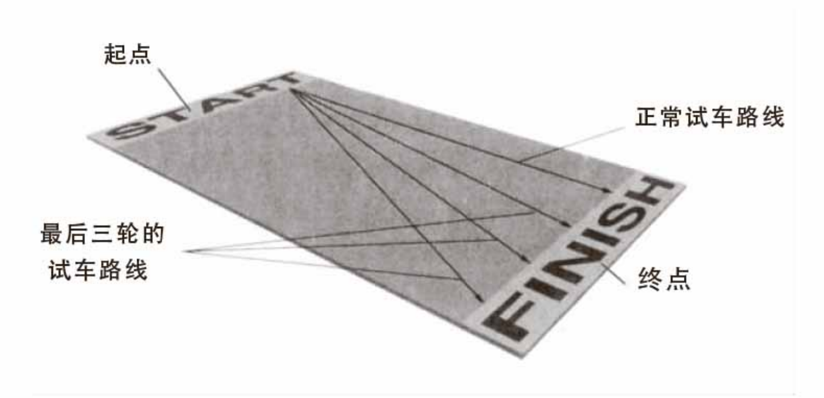



图2.5　在黄昏阳光的照射下，斯里姆在最后三轮试车中路线越偏越远。

像上面讲的，斯里姆的解释很容易理解；不过，我们在这儿重复它多少是为了下面在概念上的飞跃。南北方向和东西方向是汽车能自由活动的两个独立空间维度。（当然，它还可以在竖直方向上运动，例如爬过山坡。不过，在这儿不需要那样。）斯里姆的解释说明，即使汽车速度每回都是160千米/小时，在最后三轮里，因为它在两个方向上运动，因而在东西方向上的速度就显得比160千米/小时慢了。在前些轮试车时，那160千米/小时的速度完全都跑在东西方向上；而在最后三轮，南北方向上也有了一定的速度。

爱因斯坦发现，这种运动在两个方向上分解的思想，正是狭义相对论一切惊人的物理学事实的基础——不过我们需要明白，不光是空间维分解运动，时间维也能“分享”运动。实际上，在大多数情况下，物体运动的大部分都是在时间而不是空间度过的。我们来看这是什么意思。

在空间发生的运动，我们很小的时候就知道了。我们也知道（尽管没有这样想过），我们和我们的朋友，以及我们所有的东西，也在时间里运动。当我们抬头看挂钟的时候，或者当我们悠闲地坐着看电视的时候，钟的读数在不停地变化，不停地“在时间里向前走”。我们和周围的一切事物都会变老，不可避免地在时间里从一刻走到下一

刻。实际上，数学家闵可夫斯基（Hermann Minkowski）以及后来爱因斯坦，都倡导把时间看成宇宙的另一维——第四维——就像我们想象自己浸在三维空间一样。这听起来很抽象，但时间维的概念却是具体的。当我们想会见某个人，我们会告诉他“在空间”的哪儿见面——如第7大道53街区一个角落的大楼的9楼。那地方由三条信息（7大道，53街区，9楼）确定，是三维空间里的一个特定位置。然而，还有一点也同样重要，我们应确定在什么时候见面——如下午3点。这个信息告诉我们“在时间”的什么地方。于是，事件由4点信息来确定：3个空间的和1个时间的。这些数据就确定事件在空间和时间（或者简单地说，在时空）里的位置。在这个意义上，时间是另一维。

从这个观点说，空间与时间是全然不同的维度。那么，我们是不是还能像讲物体通过空间的速度那样来讲它通过时间的速度呢？当然可以。

能这么做的一大线索来自我们曾经遇到过的一个重要现象：当物体相对于我们在空间运动时，它的钟比我们的走得慢。就是说，它在时间里的运动速度慢了。现在来看爱因斯坦的思想飞跃，他宣布，宇宙间的一切事物总是以一个固定的速度——光速，在时空里运动。这是很奇怪的想法；我们习惯了物体运动速度远小于光速的观念，我们还反复强调这一点正是在日常生活中看不到相对论效应的原因。这都是对的。我们现在讲的是在四维——三维空间和一维时间——里的组合速度，这个推广的速度正好等于光的速度。为更彻底理解这一点，认识它的重要意义，我们还来看上面讲的那辆只有一个速度的汽车。那个速度可以在不同的维度里分解——不同的空间维和一个时间维。假如物体（相对于我们）静止不动，就是说，它不在空间运动，类似于第一轮试车，所有的运动都发生在一个维度里——不过，在现在的情形，那一维是时间。而且，相对于我们静止以及相互相对静止的所有物体，都在时间里运动——以完全相同的速度或节律衰老。然而，假如物体在空间运动，那么刚才讲的在时间的运动一定会转移一部分

到空间来。跟汽车偏离路线一样，物体运动的转移意味着它在时间里的运动比静止时慢，因为有的运动现在转移到空间里去了。就是说，当物体在空间运动时，它的钟会变慢。这正是我们以前发现的结果。现在我们看到，相对我们运动的物体的时间变慢的原因，是它在时间里的部分运动转移为空间运动了。这样，物体在空间的运动只不过反映了有多少时间里的运动发生了转移。[6]

我们还看到，这个理论框架直接包含着一个事实：物体的空间速度有一定的极限。假如物体在时间里的运动完全转移到空间来了，物体在空间的运动就达到那个最大速度。也就是说，以光速在时间里运动的物体，现在以光速在空间运动。因为所有在时间里的运动都被占有了，因此这是物体——任何物体——所能达到的最大速度。这相当于说，我们试验的汽车直接在南北方向行驶。这时候，汽车在东西方向没有留下一点儿运动；以光速在空间运动的事物，同样也没有留下一点儿在时间里的运动。因此，光不会变老；从大爆炸出来的光子在今天仍然是过去的样子。在光速下，没有时间的流逝。

**$E=mc^2$呢**

尽管爱因斯坦没有宣扬他的理论是“相对论”（他建议叫它“不变性”理论以反映光速的不变性特征），我们现在还是明白了这个词的意思。爱因斯坦的研究证明，在过去似乎分离、绝对的空间和时间的概念，实际上是相互交织的，是相对的。他还接着证明，世上的其他物理性质也是出人意外地相互关联的。他最有名的方程为我们提供了一个重要范例。在这个方程里，爱因斯坦宣布，物体的能量（E）和质量（m）不是两个独立的概念；我们可以从质量（乘以光速的平方，$c^2$）决定能量，也可以从能量（除以光速的平方）得到质量。换句话讲，能量与质量像美元与法郎一样，是可以兑换流通的。然而，与钞票兑换不同的是，这里的兑换率是光速的平方，总是固定不变的。由于这个因子很大，小质量能产生大能量。不足8.712克（0.02磅）的铀转化的能量，曾在广岛带来毁灭性的破坏；总有一天，我们

可以利用取之不竭的海水，通过核聚变获得我们世界所需要的能量。

根据这一章强调的概念，爱因斯坦方程为我们最确切地解释了一个关键问题：没有什么东西能比光更快。你可能会奇怪这是为什么。例如，我们把一个μ子用加速器加速到10.73亿千米/小时——光速的99.5%，“加把劲儿”，加到99.9%光速，然后，“真正再加把劲儿”，让它突破光速的壁垒。爱因斯坦的公式说明这样的努力是永远不会成功的。物体运动越快，它的能量越大；而根据爱因斯坦的公式我们看到，物体能量越多，它的质量越大。例如，μ子以99.9%的光速运动时，要比它静止的伙伴重得多——严格说，大约重22倍（表1.1所列的是静态粒子质量）。而物体质量越大时，把它加速就越困难。把小孩儿搭上自行车很容易，推动一辆大卡车可就是另一回事儿了。所以，当μ子越来越快时，越不容易提高它的速度。当速度为99.999%光速时，μ子的质量增加到它原来的224倍；在99.99999999%光速时，它的质量比原来大70000多倍。在速度逼近光速的过程中，质量的增加是没有极限的，因此需要无限的能量才可能使它达到或超过光速壁垒。这当然是不可能的，所以绝对不会有什么东西能比光还跑得快。

在下一章，我们会看到，这个结论也是物理学过去百年面对的第二个大冲突的根源，并最终使另一个曾令人仰慕和喜爱的理论走向死亡——那就是牛顿的万有引力理论。

注释

1.当地球那样的大质量物体存在时，并不会因为出现强大的引力使问题更复杂。因为我们关心的是水平方向而不是竖直方向的运动，所以可以忽略地球的存在。在下一章我们会彻底讨论引力作用。

2.准确地说，光在真空中的速度是10.8亿千米/小时。光在经过空气、玻璃等物质时，速度会减小，就像悬崖上落下的石头落进水里也

会减慢速度。光速的减小并不影响我们的相对论讨论，所以我们有理由完全忽略它。

3.我们为喜欢数学的读者把这些观测现象表达为定量的形式。例如，设运动的钟的速度为v，光子往返经过的时间为t秒（根据我们静止钟的观测），则当光子回到下面的镜子时，钟经过了vt的距离。现在，我们可以用

毕达哥拉斯定理来计算图 2.3 中的每条斜线距离，$\sqrt{(vt/2)^2+h^2}$，这里 $h$ 是光子钟上下镜面的间隔。于是，两条斜线的总长是 $2\sqrt{(vt/2)^2+h^2}$。因为光速是一个常数（习惯上记为 $c$），光经过这段距离的时间应该是 $2\sqrt{(vt/2)^2+h^2}/c$（秒）。这样，我们有等式 $t=2\sqrt{(vt/2)^2+h^2}/c$，解出 $t=2h/\sqrt{c^2-v^2}$。为避免混淆，我们写成 $t_{动}=2h/\sqrt{c^2-v^2}$，下标说明这个时间是我们测得的运动的钟"嘀嗒"一声经历的时间。另一方面，我们静止的钟"嘀嗒"一声的时间是 $t_{静}=2h/c$，因此简单的代数结果是，$t_{动}=t_{静}/\sqrt{1-v^2/c^2}$，它说明运动的钟"嘀嗒"一声比静止的钟需经历更长的时间。这就是说，在两个事件之间，运动的钟"嘀嗒"的次数比静止的钟少，说明运动者感觉他经历的时间更短。

4.看了下面这个实验，你会更加相信我们的结论。实验不是在粒子加速器里做的，要简单得多。1971年10月，哈费尔（J.C.Hafele，当时在圣·路易的华盛顿大学）和吉丁（Richard Keating，美国海军天文台）用铯原子钟在商务飞机上飞行了40小时。考虑了大量与引力效应（下一章讨论）有关的特征后，狭义相对论结果证明运动的原子钟经历的时间比地球上静止的同样的钟少千亿分之几秒。这就是哈费尔和吉丁的发现：运动的钟的时间真的慢了。

5.尽管图2.4正确说明了物体在运动方向上的收缩，但那图像并不是我们实际看到的样子——假如真有物体被推向光速，假如我们的眼睛或者相机能灵敏地捕捉每个瞬间！我们——或者相机——看一样东西，是收到了从那物体表面反射回来的光。但是，反射的光来自物体

不同的位置，所以我们在任何时刻看到的光经过了不同长短的路线。结果，我们看到的是一幅带着相对论视觉错乱的图像：物体不但缩短了，还旋转了。

6. 熟悉数学的读者可能知道，我们可以根据时空位置的4－矢量 $x=(ct, x_1, x_2, x_3)=(ct, \vec{x})$ 得到速度4－矢 $u=\mathrm{d}x/\mathrm{d}\tau$，这里 $\tau$ 是由 $\mathrm{d}\tau^2=\mathrm{d}t^2-c^{-2}(\mathrm{d}x_1{}^2+\mathrm{d}x_2{}^2+\mathrm{d}x_3{}^2)$ 定义的“固有时间”。于是，“在时空里运动”的速度是4－矢 $u$ 的大小，$\sqrt{(c^2\mathrm{d}t^2-\mathrm{d}\vec{x}^2)/(\mathrm{d}t^2-c^{-2}\mathrm{d}\vec{x}^2)}$，正好等于光速 $c$。现在，我们可以将等式 $c^2(\mathrm{d}t/\mathrm{d}\tau)^2-(\mathrm{d}\vec{x}/\mathrm{d}\tau)^2=c^2$ 重新写成 $c^2(\mathrm{d}\tau/\mathrm{d}t)^2+(\mathrm{d}\vec{x}/\mathrm{d}t)^2=c^2$。这说明物体空间速度 $\sqrt{(\mathrm{d}\vec{x}/\mathrm{d}t)^2}$ 的增加，一定伴随着 $\mathrm{d}\tau/\mathrm{d}t$ 的减小，后者正是物体在时间里运动的速度（物体自己的钟经历的固有时间 $\mathrm{d}\tau$ 与我们静止钟的时间 $\mathrm{d}t$ 之比）。

# 第3章　卷曲与波澜

爱因斯坦通过狭义相对论解决了关于运动的“古老的直觉”与光速不变性之间的矛盾。简单地说，我们的直觉错了——因为我们寻常的运动跟光相比太慢了，而缓慢的运动遮掩了空间和时间的真实特性。狭义相对论揭开了它们的本性，说明它们大不同于我们从前的观念。然而，修正我们对空间和时间基础的认识却不是那么轻松的事情。爱因斯坦很快就意识到，狭义相对论引发了一连串的反应，其中有一点是特别剧烈的：万物以光速为极限的概念与牛顿在17世纪后期提出的可敬的引力理论是不相容的。于是，狭义相对论在解决一个矛盾的同时，又引出另一个矛盾。经过10年艰辛甚至痛苦的研究，爱因斯坦带着他的广义相对论走出了困境。在这个理论中，爱因斯坦又一次革新了我们的空间和时间观念，他证明它们是卷曲着的，而引力就是那卷曲的波澜。

## 牛顿的引力论

I·牛顿（Isaac　Newton）1642年生在英国林肯郡。他把数学的全部力量带给了物理学的追求，改变了科学研究的面貌。他是不朽的智

者，当问题需要新的数学时，他就自己把它创造出来。约3个世纪过去后，我们才看到另一个跟他一样的科学天才。牛顿关于宇宙的行为有数不清的发现，我们在这儿关心的是他的万有引力理论。引力作用充满了我们的日常生活。它让我们和我们周围的事物安稳地站在地球的表面；它不让我们呼吸的空气逃向外层空间；它使月亮围绕着地球，把地球约束在围绕着太阳的轨道上。从小行星、行星，到恒星和星系，亿万个宇宙的精灵在永不停歇地舞蹈，引力在指挥着这台宇宙大戏的旋律。300多年来，牛顿的影响使我们理所当然地认为，这惟一的引力是天地间万物发生的根源。但在牛顿以前，没人知道从树上落下的苹果会跟围绕着太阳旋转的行星有着相同的物理学原理。牛顿大胆地迈出一步，统一了主宰天与地的物理学，指出引力是在天地间活动着的一只看不见的手。

牛顿的引力思想大概可以说是一种伟大的平均论。他认为，每一样东西绝对有一个作用于其他任何东西的引力；不论事物的物理组成如何，它总能吸引别的事物，也被别的事物所吸引。经过对开普勒（Johannes Kepler）行星运动分析的仔细研究，牛顿得到，两个物体间的引力大小仅仅依赖于两个因素：组成每个物体的物质总量和物体间的距离。所谓“物质总量”指的是构成物质的质子、中子和电子总数，它决定着物体的质量。牛顿的万有引力理论断言，物体质量越大，两个物体间的引力越大；物体质量越小，引力越小；而且，物体间距离越小，引力越大；距离越大，引力越小。

牛顿不仅定性描述了引力，还写出了定量的方程。用语言来说，方程的意思是，两个物体间的引力正比于物体质量的乘积，反比于物体间距离的平方。这个“引力定律”可以用来预言行星和彗星围绕太阳的运动，月亮绕地球的运动，火箭在太空的运动；它还更多地用来描写地球上的运动，如篮球在空气中飞行，跳水队员从跳板上旋转着跳入水池。公式预言的与实际看到的这些事物的运动惊人地一致。直到20世纪初，这些成功一直是牛顿理论不容辩驳的支柱。然而，爱因斯

坦的狭义相对论却给牛顿理论带来一个难以逾越的巨大障碍。

牛顿引力与狭义相对论不相容

狭义相对论的一个重要特征是光所限定的绝对速度。这个极限速度不仅适用于有形的物体，也适用于信号和各种形式的影响作用，认识这一点是很重要的。信息或者干扰从一个地方传到另一个地方，都不可能比光速更快。当然，比光慢的传播方式在世界上是很多的。例如，说话或者别的什么声音，是由振动以每小时1100千米的速度在空气中传播的，这与每小时10.8亿千米的光速相比确实微不足道。这两种速度的差别，在我们远离本垒观看棒球比赛时会变得很明显。当击球手击中球时，我们会先看到球被击中，然后才听到击球的声音。类似的现象发生在雷雨时。虽然闪电和雷鸣是同时发生的，但我们总是先看到闪电，后听到雷鸣。这同样反映的是光速与声速的巨大差别。狭义相对论的成功使我们知道，相反的情况——某个信号比光先到达我们——是不可能发生的。没有东西能比光更快。

问题是这样的：在牛顿引力理论中，一个物体作用在另一个物体上的引力完全决定于两个物体的质量和分开的距离，而与它们相互作用的时间无关。就是说，如果物体的质量和距离变了，则照牛顿的观点，物体将同时感觉它们之间的引力也变了。例如，牛顿理论认为，假如太阳突然爆炸了，那么1.5亿千米外的地球会立刻脱离它寻常的椭圆轨道。即使光从爆炸的太阳传到地球需要8分钟的时间，在牛顿理论中，太阳发生爆炸的消息却因为引力的突然改变而瞬间传到地球。

这个结论是直接与狭义相对论矛盾的，因为后者确信没有什么信息能比光的传播更快——瞬时传播大大地违反了相对论原理。

这样，爱因斯坦在20世纪初发现，成功的牛顿引力理论是与他的狭义相对论相矛盾的。他相信狭义相对论是正确的，尽管有数不清的

实验支持牛顿理论，他还是去寻找一种能与狭义相对论相容的引力理论。终于，他发现了广义相对论。在那个理论里，空间和时间的性质又一次经历了惊人的变革。

爱因斯坦最快乐的思想

即使在狭义相对论出现之前，牛顿的引力理论也存在一个严重的缺陷。它能高度精确地预言物体如何在引力作用下运动，却没能说明引力是什么。就是说，在物理上彼此分离（甚至分离亿万里）的物体，凭什么相互影响呢？引力是以什么方式发生作用的？这个问题牛顿本人当然也很清楚。照他自己的话讲，非生命物质不借任何其他非物质形式的中介而能无接触地相互发生作用，是无人能信的。引力也许是物质生来所固有的本性，所以一个物体能通过虚空超距地作用于另一个物体，而勿需其他任何中介作为那力的承载物和传播者。这一点在我看来真是一个伟大的谬误，我相信凡对哲学问题有足够思想能力的人都不会信它。引力必然有一个以一定规律持续作用的动因，不论这动因是物质的还是非物质的，我都留给我的读者去考虑。[11]

显然，牛顿接受了引力存在的事实，然后建立了精确描述它的作用的方程，但是没能发现它是如何产生的。他为世界写了一本引力的“用户手册”，告诉人们如何“使用”它——遵照那些指令，物理学家、天文学家和工程师们成功地把火箭送到了月球、火星和太阳系的其他行星；预言了日食和月食；预言了彗星的运动等等。但是，他留下了一个大大的谜——一个引力作用的“黑箱”，不知道那里面发生着什么。当我们玩儿CD机和个人电脑时，也处在类似的情形，我们不知道它们的内部是怎么工作的。我们只需要知道怎么用，不必知道它们怎么完成我们要它们做的事情。但是，一旦机器坏了，修理它就得靠内部运行的知识了。同样，爱因斯坦发现，虽然经过了200多年的实验证明，但狭义相对论表明，牛顿理论出现了某种难以捉摸的“破裂”，要修补它，需要完全把握引力的真正本性。

1907年的某一天，爱因斯坦坐在瑞士伯尔尼专利局办公室的桌旁，想着引力的问题。忽然，他抓住了关键的一点——经过曲折坎坷的思想历程，这一点终于把他引向一个崭新的引力理论，不仅弥合了牛顿引力的缺陷，而且彻底重构了引力的思维形式，而更重要的是，那形式是与狭义相对论完全一致的。

爱因斯坦那时想到的问题，与我们在第2章困惑过的问题有关。我们在那里强调的是，在观察者相对匀速运动的情况下，世界该是什么样子。仔细比较观察者们所看到的现象，我们发现背后藏着惊人的关于空间和时间本性的东西。但是，如果观察者是在加速运动呢？这种情况下，每个观察者看到的比泰然的匀速观察者要复杂得多，不过我们还是可以问一问，有没有什么办法来简化复杂，并将加速运动堂堂正正地带入我们新发现的空间和时间的概念？

爱因斯坦“最快乐的思想”就是那样一种简化问题的方法。为了理解他的思想，让我们走进2050年的一个故事。一天，你突然接到一个紧急电话，在华盛顿特区中心发现一颗像精心安置的炸弹模样的东西，要你去检查（你是联邦调查局首席爆破专家）。你急急赶到现场一看，果然证实了你的忧虑：那是颗核弹，威力巨大，即使埋在海底或者地壳下面，它的爆炸也会带来毁灭性的灾难。你小心翼翼地检查了它的引爆机制，没办法消除；你还发现，它像一个奇巧的饵雷，随便碰不得。炸弹装在一个刻度盘上，当盘上的数字偏离现在一半时，炸弹就爆炸。根据它的计时方式，你知道自己只有一个星期多一点儿的时间了。几百万人的命运落在你的肩上——怎么办？

看来，它在地球上的任何一个地方爆炸都是不安全的，你只有一个选择：把它送到遥远的太空去，在那儿爆炸应该不会带来什么破坏。在联邦调查局（FBI）的专案组会上，你提出这个想法，但立刻遭到了一位年轻助手的反对。“您的计划存在严重问题，”年轻的伊萨克（Isaac）告诉你，“当炸弹远离地球时，它的重量会减轻，因为地球对它的吸引力消失了。这意味着装置上刻度盘的读数会减小，炸弹

还没到安全的高度就会爆炸。”你还没来得及考虑他的意见，另一位年轻助手阿耳伯特（Albert）又站起来：“其实，细想想，还有更严重的问题，”他说，“这个问题跟伊萨克的一样重要，也许更难捉摸，请耐心听我解释一下。”你想叫阿耳伯特停下来让自己好好想想伊萨克的意见，可他总是一开口就没人堵得住。

“要把炸弹送上太空，我们只有将它绑在火箭上。火箭向上加速时，刻度盘上的读数会增大，一样会使炸弹先爆炸。原因是这样的：炸弹的基座——在那个刻度盘上，在加速的时候会比静止时更强烈地压迫它，就像坐在加速的汽车上我们的身体会向后挤压坐垫一样。炸弹‘挤压’刻度盘，就像我们挤压汽车坐垫。刻度盘受到挤压，当然会增大读数——只要偏离超过50%，它就将引爆炸弹。”

感谢阿耳伯特的解释。但是，你没有听他的话，你宁肯信伊萨克的。你沮丧地说，否定一个思想，只需要致命的一击就够了，伊萨克的意见显然是对的，确实否决了那个想法。你感觉有点儿绝望了，问大家还有没有新的建议。这时，阿耳伯特有一个漂亮想法。“关于第二点，”他接着说，“我想您的看法还没有完全绝望。伊萨克说的，炸弹装置升入太空时，地球引力会消失，就是说，刻度盘的读数会减小。而在我看来，火箭向上的加速度会使炸弹挤压刻度盘，就是说，盘的读数会增大。两种观点放在一起，我们发现，如果在每一时刻精确调整火箭向上的加速度，两种效应就会彼此抵消！具体说，在升空的初始阶段，火箭还能完全感觉地球的引力，这时候加速度可以不那么大，我们还能在那50%的空隙里。火箭离地球越来越远，感觉的引力越来越小，这时候我们需要增大加速度来克服引力的不足。因为上升加速度导致的读数增加，与引力消失导致的读数减小，可以完全抵消，这样，我们实际上能保证刻度盘上的读数一点儿也不改变！”

你慢慢发现阿耳伯特的建议有点儿意思。“换句话讲，”你回答说，“向上的加速度可以替代引力。我们能以适当的加速运动来模拟引力效应。”

“完全正确。”阿尔伯特答应。

“那么，”你接着说，“我们可以把核弹弄到太空去，通过精心调节火箭加速度，还可以确保刻度盘上的读数不会改变。这样，在地球的安全距离以内就不会发生爆炸了。”于是，你可以利用21世纪的火箭技术来协调引力和加速运动，从而避免一场灾难。

引力与加速运动密切关联着，正是爱因斯坦在一个快乐的日子在伯尔尼专利局的办公室里想到的最关键的一点。虽然核弹历险说明了这一思想的基本特征，但我们还是应该用接近第2章的方法再把它重复一遍。先回想一下，在封闭的没有窗户、没有加速的列车车厢里，我们无法确定自己的速度。不论速度多大，车厢看起来都是一样的，在车厢里做的实验也得出同样结果。从更基本的意义说，如果没有外面的路标做参考，我们不能以速度来定义某个运动状态。另一方面，如果列车是加速运动的，即使在封闭的车厢里，我们也能感觉到有力作用在身体上。例如，你坐在加速的列车上，面对着前进方向，你会感觉坐椅有股力量作用在背上，跟阿耳伯特讲的汽车的情形一样。同样，假如列车向上加速，你会感到地板作用在脚上的力量。爱因斯坦发现，在小小的车厢里，我们不能区别加速的情形与没有加速而有引力的情形：如果大小调节适当，来自引力场的力与来自加速运动的力是不可能区分的。假如车厢静静地停在地面上，我们的脚下会感受那熟悉的来自地板的力，就仿佛车厢在向上加速；阿耳伯特在探索如何把恐怖的核弹送进太空时，考虑的也是这种等效性。假如车厢向后倒下来停在地上，我们的后背会感觉坐椅的力量（使我们不致落下），与列车水平向前加速时的感觉一样。爱因斯坦将加速运动与引力的不可分辨的性质称作等效原理。它在广义相对论里起着核心的作用。[1]

我们将看到，从狭义相对论开始的工作，由广义相对论完成了。狭义相对论通过相对性原理确立了不同观察者的观点都是平等的；物理学定律对一切匀速运动的观察者都是一样的。但这是有限的平等，它排除了数不清的其他观点——那些加速运动者的观点。现在，爱因

斯坦在1907年的发现告诉我们如何将所有的观点——匀速的和加速的——纳入一个平等的构架。在加速的无引力场的观点与非加速的有引力场的观点之间不存在任何差别，所以我们可以借后一个观点说，所有的观察者，不论运动状态如何，都可以认为自己是静止的而“世界的其他事物在他们身边运动”，不过，在他们周围出现了某个引力场。在这个意义上，广义相对论通过引力保证所有可能的观察者的观点都一样站得住脚。（以后我们会看到，第2章讲的因为加速运动而出现的两个人之间的区别——在乔治打开喷气包追赶格蕾茜时，会变得比她年轻——也可以不用加速度而用引力来说明。）

引力与加速运动的这种深层联系当然是惊人的发现，但爱因斯坦为什么为它感到快乐呢？简单地说，引力太神秘了，尽管充满了无边的宇宙，却令人难以捉摸。另一方面，加速运动虽然比匀速运动复杂一些，却是具体而实际的。爱因斯坦发现了两者的基本联系，意识到他可以靠对运动的认识去获得对引力的理解。即使凭爱因斯坦的天才，实现这个计划也不是那么容易的。不过，这个方法最终还是结出了广义相对论的硕果。为了那个目标，爱因斯坦还建立了统一引力与加速运动的第二种联系：空间和时间的弯曲。现在我们就来看它。

加速度与时空卷曲

引力问题几乎令爱因斯坦着魔了。在伯尔尼专利局办公室冒出那个“快乐的思想”大约5年以后，他写信告诉物理学家A·索末菲（Arnold Sommerfield），“我现在完全被引力问题占有了……有一点是肯定的——我生来还从未有过什么事情这样困扰着我……与这个问题相比，原先的[狭义]相对论不过是一场儿戏。”[12]

1912年，他又迈出了关键的一步。他用狭义相对论来联结引力和加速运动，得到一个虽然简单却很微妙的结果。为跟上他的论证，最简单的办法是像他做的那样，考虑一种特殊的加速运动。[13]回想一下，物体的加速指的是要么改变速度大小，要么改变运动方向。为简

单起见，我们考虑只改变物体运动方向而速度大小保持固定的加速运动。特别地，我们考虑圆周上的运动，这种运动可以在游乐园的“龙卷风”转盘上亲身体验。假如你害怕自己的身体受不了那样的折腾，紧紧地背靠着高速飞旋的玻璃纤维环的内壁，你会像经历别的加速运动一样，觉得自己像要被径向地抛出去，而环壁在紧紧地顶着你的背，你在圆环上一点儿也动不了。（其实，高速的旋转会把你牢牢地“钉”在玻璃纤维的环上，即使脚下空了，你也不会滑落下去。当然，那跟这儿的讨论无关。）假如环非常光滑，你闭上眼睛，几乎会感觉自己正躺在床上——环壁对你背的压力就像床在支撑着你。我们说“几乎”，是因为你还能感觉到寻常的“向下”的重力，头还没有完全“转晕”。不过，如果那转环是在太空，还是转那么快，你真会感觉自己是躺在家里的床上。另外，假如你想“起床”来沿着玻璃纤维的环散散步，你会感觉双脚仿佛踏在家里的地板上。实际上，太空站就是设计成这样旋转的，让你能在太空中感觉“故乡”的引力。

我们跟着爱因斯坦用旋转的环的加速度来模拟引力，现在可以来看环里的人所感觉的空间和时间是什么样的。以我们的例子来说，爱因斯坦的论证是这样的：我们静止的观察者很容易测量转环的周长和半径。例如，为了测周长，我们可以仔细地贴着转环用尺子一步一步地量；为测半径，我们可以用同样的办法，将尺子从转轴那一点一节节摆到环的边缘。我们发现，周长与半径之比是π的2倍，约6.28——与画在纸上的任何圆圈一样，这是我们在中学几何里学过的。但是，在转环上的人会看到什么样的情形呢？

还是让斯里姆和吉姆来告诉我们吧。这会儿，兄弟俩正在转环上玩儿呢。我们请两人各拿一把尺子，斯里姆测量周长，吉姆测量半径。为看得更清楚些，我们来鸟瞰一下那个转环，如图3.1。在图中我们画了一个箭头，说明在那个时刻各点的运动方向。当斯里姆开始测量周长时，我们从旁发现他将得到不同的结果。他把尺子贴着环一节节测量时，我们会看到尺子缩短了，这不过是第2章讨论过的洛伦

兹收缩，即物体的长度沿运动方向缩短。既然尺子缩短了，他必须多测量几步才能测完整个周长。而他自己还以为尺子仍然是30厘米（因为斯里姆与尺子间没有相对运动，所以他觉得尺子的长度跟平常一样），所以他测得的周长比我们测的更长。（如果你觉得奇怪，可以看看后面的注释。[2]）



图3.1　斯里姆的尺子沿着转环运动的方向，长度缩短了。吉姆的尺子在径向支架上，与运动方向垂直，所以长度没有缩短。

那么，半径呢？当然，吉姆也是用尺子一节节去测量转环的径向支架的长度，从我们的眼睛看，他测的长度跟我们相同。原因是，他的尺子并没有（像测量周长那样）指向每一瞬间的旋转方向。实际上，尺子是垂直于运动方向的，所以不会发生长度的收缩。于是，吉姆得到的径向长度跟我们是完全一样的。

但是，当斯里姆和吉姆计算周长与半径之比时，他们会得到一个

比我们的2π更大的数，因为这时的周长大了，而半径是一样的。这可真是奇怪。一个圆的东西，怎么可能违反古老的法则呢——对每个圆来说，那个比值不都应该是2π吗？

爱因斯坦是这样解释的：古希腊发现的那个法则只对平面上的圆才成立。我们知道游乐园里哈哈镜凹凸的镜面会扭曲人的面目，同样，如果把圆画在卷曲的面上，寻常的空间关系也会被扭曲：周长与半径之比往往不等于2π。

我们来比较一下图3.2中的三个半径相同的圆。注意，它们的周长是不同的。图3.2（b）画在球面上的圆的周长比图3.2（a）画在平面上的圆的周长小，尽管它们的半径是一样的。弯曲的球面使圆的径向直线慢慢聚合，结果周长变短了。在图3.2（c）中，圆仍然画在曲面上——在马鞍面上，但它的周长却比平面的圆长；马鞍形的弯曲特点是使圆的径向直线慢慢散开，从而使周长增大了。这些事实意味着，周长与半径之比，在图3.2（b）小于2π，在图3.2（a）等于2π，而在图3.2（c）大于2π。比值与2π的偏离，特别是图3.2（c）的情形，正是我们在转环的例子中看到的。根据这个发现，爱因斯坦提出空间弯曲的概念，以解释为什么“正常的”欧几里得几何被破坏了。千百年来人们在儿童时代学习的古希腊人的平面几何，根本不适用于转环上的人，我们需要用图3.2（c）示意的那种更一般的弯曲空间的几何来代替它。[3]

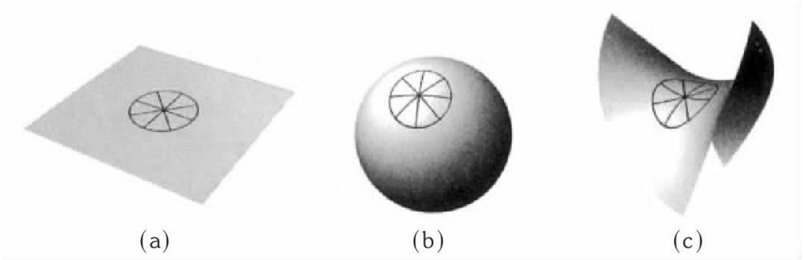



图3.2　球面上的圆（b）的周长比平面上（a）的更短，而马鞍面上的圆（c）的周长更长，尽

管三个圆的半径是一样的。

就这样，爱因斯坦认识到，我们熟悉的被古希腊人奉为法则的空间几何关系——那些与平直的空间图像（如桌面上的圆）相伴的关系，在加速运动的观察者眼里是不成立的。当然，我们只讨论了一种特殊的加速运动；但爱因斯坦证明了，在所有加速运动的情形，空间都是弯曲的。

实际上，加速运动不光导致空间弯曲，也导致类似的时间弯曲。（历史上，爱因斯坦先关注的是时间弯曲，然后才发现空间弯曲的重要性。[3]）坦白说，我们并不奇怪时间也会弯曲，因为我们已经在第2章看到狭义相对论明确地把空间和时间统一起来了。这种统一，闵可夫斯基曾在1908年的一次演讲中以诗一般的语言作了概括："从今往后，空间也好，时间也好，都将躲进阴影，只有两者的某种统一才能独立地存在。"[14]用更普通（不过也很不精确）的话来说，狭义相对论将空间和时间编织到一个统一的时空结构里，向我们宣布"凡对空间正确的，对时间也正确"。但问题跟着来了：弯曲的空间可以用卷曲的图形来表现，那弯曲的时间是什么呢？

为回答这个问题，我们还是把它交给转环上的斯里姆和吉姆，请他们做一个实验。斯里姆背靠着环站在径向支架的一端，吉姆从旋转轴心沿着支架慢慢向他爬过去。吉姆每爬几步就停下来，与斯里姆对一下表。他们发现了什么呢？从我们静止的旁观者看，还是那个结论：两人的表不同步。这个结果的原因在于，我们看到斯里姆和吉姆在以不同的速度运动——在转环上，离轴心越远，转过的距离越长，因此旋转的速度越快。但根据狭义相对论，你动得越快，你的表走得越慢。于是，我们发现斯里姆的表比吉姆的慢。而且，两人还会发现，在吉姆爬向斯里姆的过程中，他的表越走越慢，越来越接近斯里姆的表。这反映了一个事实：当吉姆在支架上越爬越远，他的旋转速度越来越接近斯里姆。

我们的结论是，对于转环上的观察者（如斯里姆和吉姆）来说，

时间的速度依赖于各人的确切位置——在这里，即他们离中心的距离。这说明了我们讲的弯曲时间：假如时间在不同的位置上有不同的速度，我们就说时间是弯曲的。对我们现在的讨论，还有特别重要的一点，吉姆在向外爬的时候会注意到另一件事情。他将感觉一股强大的力量把他向外推，因为他离中心越远，不但速度增加了，加速度也大了。于是，我们看到，在旋转的环上，大的加速度是与缓慢的钟联在一起的——就是说，加速度越大，时间弯曲越强烈。

爱因斯坦靠这些发现迈出了最后一步。他已经证明了引力与加速运动在现象上是不可分辨的，现在他又发现加速运动联系着空间和时间的弯曲，接下来他揭开了引力“黑箱”的秘密——引力是以什么机制发生作用的。据爱因斯坦的观点，引力就是空间和时间的弯曲。这是什么意思呢？

广义相对论基础

为理解这种新的引力观，我们考虑实际的行星绕恒星运动的情形，例如，地球绕太阳运行的情形。在牛顿的引力论里，太阳把地球限制在轨道上，靠的是一根“看不见的绳子”，那根引力的“绳子”仿佛从太阳生出来，瞬间穿过遥远的空间距离，把地球套住（当然，地球也同样一下子抓住了太阳）。实际发生了什么，爱因斯坦提出了新的概念。为了讨论爱因斯坦的方法，我们最好能有一个容易把握的具体形象的模型。那样可以从两个方面将问题简化。第一，我们先不管时间，只关心空间的视觉模型，然后再把时间包括进来讨论。第二，为了让图像能在纸上表现出来，我们将经常用二维的类比来替代三维的空间。从考虑这样的低维模型得到的大多数结果，都可以直接用于三维的物理空间。因此，简单的模型是有力的思维方式。

在图3.3中，我们运用了这种简化方式，把我们宇宙的空间画成一个二维的区域。图中的网格不过用来确定位置，就像我们以街道网来确定城市里的位置一样。当然，我们说城市的某个地址，往往要确

定它在二维街道网上的位置，还要说明它在竖直方向上的位置，例如在几楼几号。为了让图像更简洁，我们在二维类比的图中压缩了第3个空间方向上的东西。

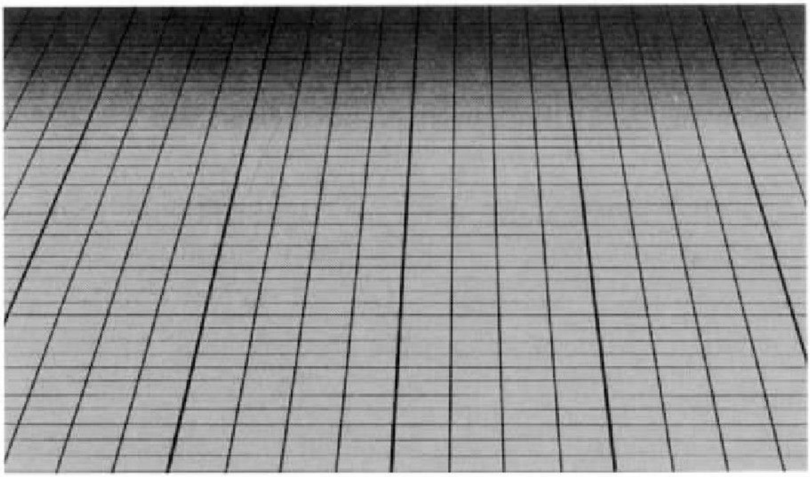


图3.3　平直空间示意图。

爱因斯坦猜想，当没有任何物质或能量存在时，空间应该是平直的。用二维模型来说，空间的“形状”应该像一张光滑的桌面，如图3.3。这也是几千年来人们普遍怀有的我们宇宙的空间图像。那么，假如空间出现一个大质量物体（如太阳），会发生什么事情呢？在爱因斯坦之前，人们会说，什么也不会发生，他们认为，空间（和时间）不过是一个死的剧场，为宇宙提供一个表现自己的舞台。但是，跟着爱因斯坦的思路，我们将走向一个不同的结论。

像太阳那样的大质量物体（实际上，任何物体）对其他物体都有引力作用。在那个可怕的“核弹事件”里，我们知道了引力与加速运动是不可分辨的。在转环游戏里，我们知道描写加速运动需要弯曲空间的关系。引力、加速运动与弯曲空间的联系启发爱因斯坦提出一个惊人的观点：物质（如太阳）的存在导致它周围的空间结构发生弯曲，如图3.4。这是我们常看到的一幅图，像一张橡皮膜上放一只保龄

球，空间结构因大质量物体的存在而发生扭曲。照这个不同寻常的看法，空间不再仅仅是被动的宇宙活动的舞台；空间的形状倒是环境事物的反映。

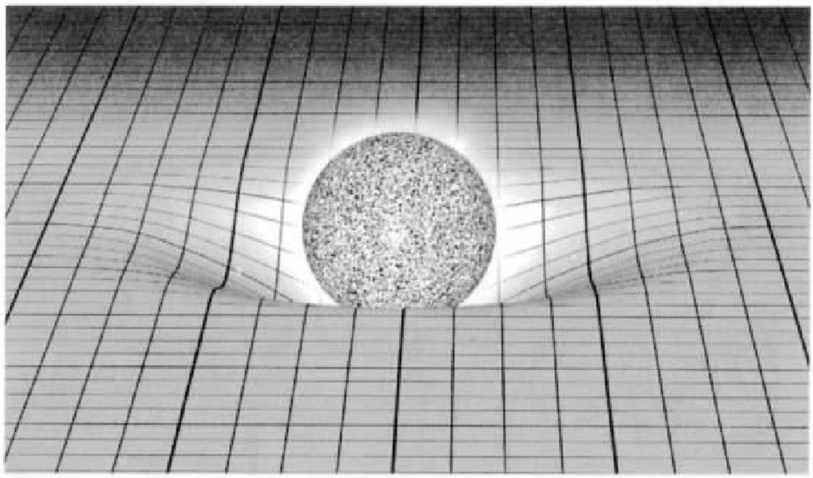


图3.4　大质量物体使空间结构发生弯曲，就像一只保龄球放在橡皮膜上。

另一方面，当太阳附近的物体经过空间扭曲的结构时，扭曲的空间也会影响它们的运动。用保龄球和橡皮膜的类比来讲，假如我们以一定的初始速度在膜上放一粒小滚珠，则它滚动的路线依赖于膜中间有没有球。如果没有球，膜还是平坦的，小珠子会沿一条直线滚过去。如果有球，膜被扭曲了，小珠子将沿着曲线滚动。实际上，如果忽略摩擦，我们可以让小珠子以适当的速度和方向滚动，它可以沿一条回归的曲线绕着中间的球滚动——就是说，“它滚进了轨道”。显然，这个例子可以用来说明引力。

太阳就像那只保龄球，它使周围的空间结构发生弯曲，地球就像那颗滚珠，被弯曲了的空间卷入它的轨道。只要速度的大小和方向适当，地球也会像滚珠那样绕着太阳转动。地球运动所受的这种影响，就是我们通常所说的太阳对地球的引力作用，如图3.5。现在我们看到，爱因斯坦不同于牛顿的是，他确定了引力传播的机理：空间的弯

曲。在爱因斯坦看来，把地球“绑”在轨道上的“引力绳”，并不是太阳的神秘的瞬间作用，而是因为太阳的存在所导致的空间的弯曲。

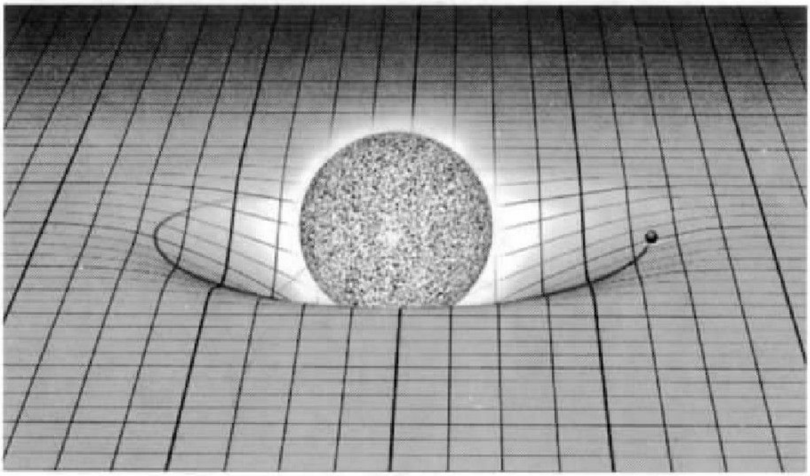


图3.5　地球在绕着太阳的轨道上运行，是因为地球滚入了弯曲空间的一道“沟谷”。更准确地说，它走的是在太阳周围弯曲区域里“阻力最小”的路线。

这幅图景帮助我们以新的方法认识了引力的两个基本特征。第一，在爱因斯坦的引力图像中，物体质量越大，所导致的空间扭曲越强，就像保龄球越大，橡皮膜的扭曲也越大。这意味着物体质量越大，它能作用于其他物体的引力就越大，这跟我们的经验是一致的。第二，距保龄球越远的地方，那里的膜的变形越小；同样，距大质量物体越远，空间的弯曲越弱。这也是我们熟悉的引力特性：物体相距越远，引力作用越小。

还有一点很重要，那就是小滚珠也会使橡皮膜弯曲，尽管那弯曲很小。同样，地球作为一个有质量的物体，当然也能使空间结构发生弯曲，不过比太阳的小得多。用广义相对论的话讲，地球就是这样带着月亮在轨道上运行的，我们也是因为这一点才能站在大地上。当跳伞者从天空落下时，他是在沿着地球质量产生的弯曲的空间结构向下滑行。另外，我们每一个人也跟其他有质量的物体一样，能使我们身

体近旁的空间结构发生弯曲。当然，我们小小的身体只能引起一点小小的波动。

总的说来，爱因斯坦完全同意牛顿说的“引力必然有一个动因”，而且响应了牛顿的挑战，考虑了他“留给我的读者去考虑”的问题。根据爱因斯坦的理论，引力的动因是宇宙的结构。

几个缺点

橡皮膜与保龄球的例子，让我们具体形象地把握了我们所说的宇宙空间结构的弯曲是什么意思。物理学家常用类似的比喻来帮助自己更直观地认识引力和弯曲。然而，尽管这办法很有用，膜与球的类比还是不够完美，为把问题讲清楚，我们来看它有哪些缺点。

第一，当太阳使它周围的空间发生弯曲时，并不像保龄球那样是因为它被引力“拉下来”；在保龄球的例子中，是因为地球的引力作用才使膜发生弯曲的。对太阳来说，没有别的什么东西在“拉它”。相反，爱因斯坦告诉我们，空间的弯曲才是引力。只要有物质的存在，空间就会发生弯曲。同样，地球也不像弯曲膜上的那颗小滚珠，它在轨道上运行并不是因为有什么别的外来的东西把它引入弯曲空间里的沟谷。事实上，爱因斯坦证明，物体在空间（准确说是时空）的运动总沿着可能的最短路线——“可能的最容易的路线”或“阻力最小的路线”。如果空间是弯曲的，这样的路线也会弯曲。所以，球与膜的类比尽管让我们直观看到了大质量物体（如太阳）如何扭曲空间，又如何影响其他物体（如地球）的运动，但空间扭曲的物理学机制却完全不是那样的。球与膜的模型所依据的是我们在传统的牛顿框架内对引力的直观认识，而弯曲空间的却是爱因斯坦重构的引力框架。

球与膜类比的第二个缺点在于橡皮膜是二维的。实际情况是，太阳（以及一切有质量的物体）扭曲了它周围的三维空间，当然这是很难用图像来表达的。图3.6试着表现了一下。太阳周围的所有空间

——它的“下面”、“旁边”和“上头”——都经历了相同性质的扭曲变形，图3.6只画出了一部分。物体（如地球）就在这样的弯曲的空间环境里穿行。也许有人觉得奇怪——地球为什么不会闯入图中的那道“墙”呢？可是别忘了，空间不像橡皮膜，不是摸得着的壁垒。图中的弯曲网格不过是从弯曲的三维空间里剪下来的两张薄片，而你和我、和地球以及其他万事万物一样，都浸没在那个三维空间里，在其中自由地活动。也许你会觉得这把问题说得更令人困惑了：如果我们真的浸没在空间的结构里，为什么没有感到它的存在呢？其实，我们感觉到了。我们感觉了引力，而空间正是引力发生作用的中介。大物理学家惠勒（J.Wheeler）常常这样描述引力：“质量牵着空间，告诉空间如何弯曲；空间牵着质量，告诉质量如何运动。”[15]

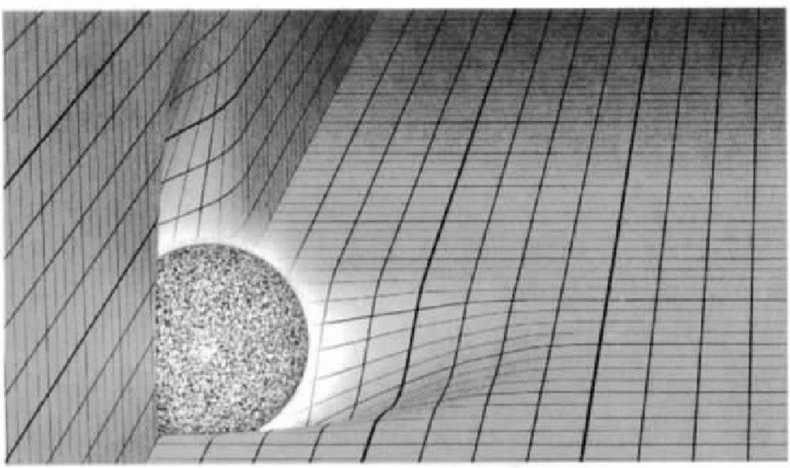


图3.6　太阳周围三维空间的弯曲。

类比的第三个缺点是我们把时间维压缩了。这样做原是为了使图像更清晰；尽管狭义相对论明确指出我们应该像3个空间维那样来思考时间维，但时间却总是难得“看见的”。不过，转环的经历说明，加速度既弯曲了空间，也弯曲了时间，引力当然也该如此。（实际上，广义相对论的数学证明，在像地球围绕太阳运行的这种相对缓慢的运

动中，时间的弯曲对地球运动的影响要远远小于空间的弯曲。）下一节过后，我们还会回来讨论时间的弯曲。

上面讲的三个缺点是很严重的，但是只要我们在心里把握了它们，球与膜所表现的弯曲空间的图景还是满可以借来形象地概括爱因斯坦的新引力观。

矛盾解决了

爱因斯坦描绘了一个清晰的引力作用图景，空间和时间在那里也成了动力学的参与者。不过，关键的问题是，这个新建的引力理论的框架能不能解决困扰着牛顿引力理论与狭义相对论的矛盾？回答是，它真解决了这个矛盾。我们还是借膜的类比来说明基本的思想。我们想象，在没有保龄球的时候，有一颗珠子沿着平坦的膜上的一条直线滚动。当我们把球放上膜时，小珠子的运动会受影响，但那不会在瞬间发生。如果把这过程拍摄下来，看它的慢动作，我们会看到，保龄球引起的扰动像水池里的波纹一样向周围扩散开去，然后到达滚珠的位置。不久，膜面的波动平息下来，我们看见一张静止的弯曲了的膜。

空间结构也是这样的。没有质量存在时，空间是平直的，小质量物体可以安然地静止其间或者匀速地运动。当大质量物体出现时，空间会弯曲——但像膜一样，不会瞬间地扭曲；扰动从大物体开始，然后向外扩张，最后形成弯曲的空间结构，传达那个庞然大物的引力作用。在我们的类比中，扰动在膜面上向外的扩张速度由膜的组成材料决定。在广义相对论的情形，爱因斯坦可以计算宇宙空间结构的扰动以多大速度传播，他发现那正好是光速。我们在前面曾讨论过一个假想的例子，太阳熄灭了，通过引力的改变影响地球——现在我们知道，那影响不会瞬间传到地球。实际上，当物体改变位置或甚至被风吹动时，都会使时空结构的扭曲形态产生扰动，扰动以光速向外扩张，这满足了狭义相对论以光速为宇宙极限速度的要求。所以，在太

阳爆炸8分钟以后，地球上的我们才知道太阳熄灭了，而在这同一时刻我们也感到太阳的引力消失了。爱因斯坦的新理论就这样解决了矛盾；引力扰动与光同步，却总也超不过它。

再谈时间弯曲

图3.2、图3.4和图3.6基本上让我们看到了什么是“弯曲的空间”，那就是空间形态被扭曲了。物理学家曾用类似的图像来表现“弯曲的时间”，但那太难说明白了，所以我们不画那些图。我们还是学斯里姆和吉姆坐转环的例子来体验引力产生的时间弯曲。

为此，我们再来看看格蕾茜和乔治。这回他们不在漆黑的太空，而是漂浮在太阳系的边缘。他们的太空服上还戴着巨大的数字钟，是原来校准好了的。为简单起见，我们忽略行星的影响，只考虑太阳引力场的作用。在乔治和格蕾茜附近泊着一艘飞船，从飞船放下一根长线，伸向太阳的表面。乔治顺着那线慢慢接近太阳。每过一定时间，他就停下来，与格蕾茜比较他们的钟走过的时间。广义相对论的时间弯曲的预言说明，乔治的钟将比格蕾茜的钟越走越慢，因为他经历的引力场越来越强。就是说，他离太阳越近，他的钟越慢。从这个意义说，引力像扭曲空间那样也扭曲了时间。

应该看到，这里的情形与第2章不同。在那里，乔治和格蕾茜是在虚空里以不变的速度相对运动着，而现在两人不再有那样的对称地位。与格蕾茜不同的是，乔治感到引力在越变越强——他离太阳越近，就得费更大的气力抓紧那根线才不会被太阳的引力拉下去。在格蕾茜看来，乔治的钟慢了；乔治自己也同意他的钟慢了。在这里，两个角色的地位是不“平等的”，也就不会有以前遇到过的相反的结论。事实上，这正是第2章里乔治打开喷气袋去追赶格蕾茜所发生的事情。乔治的加速度使他的钟走得肯定比格蕾茜的慢。我们现在知道，经历加速运动与感觉引力作用是一回事，乔治现在的情形也满足这个原理，所以我们看到，他的钟和发生在他生命里的一切，都比格蕾茜

的慢。

在一颗普通恒星（如太阳）的表面，引力场对时间的影响是很小的。例如，让格蕾茜停在离太阳10亿千米以外，而乔治下到离太阳表面几千米的地方，他的钟的节律大约为格蕾茜的99.9998%，的确慢了，但慢得不多。[4]但是，假如乔治顺着长线爬到一颗中子星的表面，他的钟的节律将是格蕾茜的76%。虽然中子星质量与太阳差不多，但密度却是太阳的千万亿倍，所以它的引力场比太阳强得多。更强的引力场，如黑洞的外面（下面讨论），将使时间走得更慢；引力场越强，时间弯曲越严重。

广义相对论的实验验证

大多数研究广义相对论的人都会醉心于它的美妙。牛顿那冷冰冰的、机械的空间、时间和引力的概念，被爱因斯坦以一种动力学的、几何的弯曲的时空图景取代了，引力嵌入了宇宙的基本结构。在最基本的水平上，引力成为宇宙的主要部分，而勿需添加多余的结构。空间和时间卷曲着、褶皱着、激波荡漾着，活生生地表现着我们所说的引力。

不管有多美，物理学理论最终还得靠它精确解释和预言物理学现象的能力来检验。牛顿的引力理论自17世纪末出现以来直到20世纪初，总是一路旌旗，经历了无数考验。不论是从斜塔落下的铁球、向太空飞去的火箭，还是在太阳身边往来的彗星，所有现象都能从牛顿理论得到非常精确的解释；牛顿理论提出的预言也在不同条件下得到了数不清的证实。我们说过，向这样一个在实验上无比成功的理论提出疑问，只是因为它那瞬时传递引力的性质与狭义相对论矛盾。

在我们生存的低速世界里，狭义相对论的效应是极端微弱的，尽管它们对认识空间、时间和运动起着关键作用。同样，与狭义相对论相容的广义相对论与牛顿引力理论的差别，在大多数普通情形也都小

得可怜。这是好事，也是坏事。好的方面在于，如果哪个理论想来取代牛顿的引力理论，它在牛顿理论经过检验的场合应该拿出更好的结果；不好的地方是，我们很难从实验判别哪个理论更好。区别牛顿理论与爱因斯坦理论需要极高的实验测量精度，而实验却敏感地依赖于两个理论是如何不同的。例如，扔出一只球，用牛顿和爱因斯坦的引力理论来预言球会落到什么地方，两家结果会不一样，但差别很小，超出了我们通常实验检验的能力。我们需要更精巧的实验，爱因斯坦曾经提出过一种。[5]

我们都是在晚上看星星，当然它们白天也在；但我们白天往往看不见，因为遥远微弱的星光总被太阳的光芒掩盖了。不过，日食的时候，月亮遮住了太阳，遥远的星星也看得见了。但太阳还是在发生影响。来自遥远恒星的光在到达地球的路上一定会从太阳近旁经过，爱因斯坦的广义相对论预言太阳使它周围的空间和时间发生弯曲，这弯曲将影响星光的路线。来自远方的光子在宇宙的结构里穿行，结构弯曲了，光子的运动当然也会受影响，这与有质量的物体没有什么不同。从太阳掠过的光信号在到达地球时，路线已经历了最大的偏转。日食使我们有机会看到那些没有被淹没的掠过太阳的星光。

光线偏转的角度可以简单测量。偏转的星光带来的错觉是，我们看到的恒星的位置移动了。将我们在日食看到的恒星位置与半年前（或后）当地球在轨道另一端时我们在夜晚（没有太阳弯曲的影响）看到的恒星的实际位置相比较，就能准确测量位置偏移了多少。1915年11月，爱因斯坦用他的新引力观点计算了星光掠过太阳应该偏转的角度，结果是0.00049度（1.75弧秒，1弧秒等于1/3600度）。这个小小的角度与从3千米外看一枚硬币张开的角度一样。不过，那时的技术已经能够测量这么小的角度。1919年5月29日日食期间，在格林尼治天文台台长F·戴森爵士（Sir Frank Dyson）的激励下，英格兰皇家天文学会秘书、著名天文学家A·爱丁顿（Arthur Eddington）组织了一只考察队到西非海岸的普林西比岛去检验爱因斯坦的预言。

普林西比的日食照片（还有戴维森（Charles　Davidson）和克罗梅林（Andrew　Crommelin）率领的另一只英国考察队从巴西索布雷尔拍回的照片）经过5个月的分析后，1919年11月6日，皇家学会和皇家天文学会联合举行会议宣布，爱因斯坦基于广义相对论的预言得到了证实。从前的空间和时间概念被彻底推翻了——这个胜利的消息很快就超越了物理学的小圈子，爱因斯坦成了举世闻名的人物。在第二天的伦敦《泰晤士报》上，我们看到这样的大标题："科学革命——宇宙新理论——牛顿理论大崩溃。"[16]这是爱因斯坦辉煌的一刻。

接下来的几年里，人们仔细审查了爱丁顿关于广义相对论的证据。测量中有许多困难和不确定的因素，很难对原来结果的可靠性提出什么疑问。不过，近40年来，利用新的技术进步进行的大量实验以极高的精度检验了广义相对论的诸多方面，所有预言都被证实了。人们不再怀疑，爱因斯坦的引力图像不但与狭义相对论相容，而且还提出了比牛顿理论更接近实验结果的预言。

黑洞、大爆炸和空间膨胀

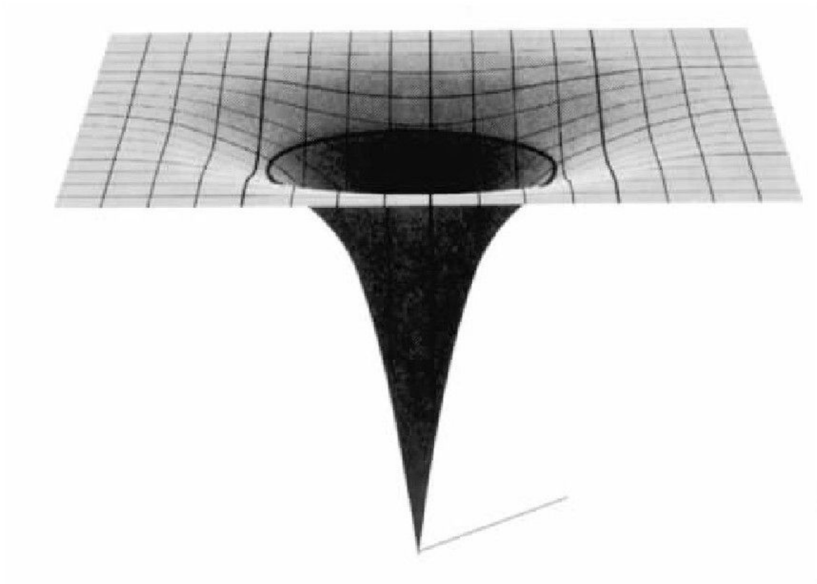
狭义相对论在高速运动的情形会显著地表现出来；广义相对论则在物体质量大、空间和时间相应地弯曲剧烈时发挥自己的作用。我们来看两个例子。

第一个例子是德国天文学家K·史瓦西（Karl Schwarzschild）发现的。那是1916年第一次世界大战期间，他正在俄国前线，一面计算他的弹道曲线，一面学习爱因斯坦的引力新发现。令人惊讶的是，爱因斯坦完成广义相对论还没几个月，史瓦西就用这个理论得到一幅完整而精确的图像，描绘了完全球状星体附近的空间和时间是如何弯曲的。他把结果从俄国前线寄给爱因斯坦，爱因斯坦代表他向普鲁士科学院做了报告。

史瓦西的研究——我们现在知道的"史瓦西解"——不仅从数学上

证实并精确化了图3.5示意的空间弯曲，而且揭示了广义相对论的一个奇妙结果。他证明，假如星体质量聚集在一个足够小的球状区域，质量除以半径超过某个特别的临界值，那么时空将产生剧烈的卷曲，包括光在内的一切事物都将卷缩在星体附近，而不可能逃脱它的引力的掌握。因为连光都跑不出这种“压缩的星体”，所以它们起初叫黑星或冻星。更动听的名字是多年以后惠勒发明的，他称它们为黑洞——因为不发光，所以“黑”；因为靠近它们的任何东西都会落下去，一去不回，所以是“洞”。这个名字一直叫到今天。

图3.7表现了史瓦西的解。尽管黑洞很“贪婪”，但如果物体从“安全的”距离经过它时，也就像经过一颗普通的恒星；虽然有一定的偏转，却还能继续它的快乐旅行。但是，不论什么材料构成的物体，只要离黑洞太近——近到所谓黑洞的事件视界以内——它就完了；它将不可抗拒地被拉向黑洞的中心，去忍受那无限增大的最终毁灭一切的强大引力。假如你的脚先落进了事件视界，那么，当你向黑洞中心逼近时，你会感觉越来越难受。黑洞引力将大得吓人，作用在你脚上的力会比头上的大得多（因为脚先落进黑洞，它离中心比头更近一点儿）；多大呢？它会把你拉长，然后把你的身体撕成碎片。



黑洞的中心

图3.7　黑洞令周围时空结构发生严重卷曲，任何物体落进它的“事件视界”——图中的黑圆圈——都不可能逃脱它引力的掌握。没人完全知道黑洞最深处的一点在发生着什么。

反过来看，假如你有先见之明，在黑洞附近游荡时万分小心，不敢越过视界半步，那么你可以借助黑洞去经历一次奇遇。例如，你找到了一个比太阳重1000倍的黑洞，想凭着一根长线，像乔治接近太阳那样爬到黑洞视界上面3厘米的地方。我们说过，引力场导致时间弯曲，这意味着时间经历会慢下来。实际上，因为黑洞的引力场太强了，所以你的时间经历不但会慢，简直要慢到家了。当你在地球上的朋友们的钟响过1万次，你的钟可能才响1次。如果你这样在视界上漂浮1年，然后沿着长线向上爬回等着你的飞船，然后经过短暂而愉快的旅行回到地球的家。当你踏上地球，你会发现距你当初离开已经过了1万年！这样，黑洞成了某种时间机器，让你能走到地球遥远的未来。

现在我们具体来看有关的几个极端的数字。如果太阳质量的恒星成了黑洞，它的半径不会是现在的大小（约70万千米），而将不足3

千米——想想看，那就是说可以把太阳拿到曼哈顿岛的一个角落。一小勺这样挤压过的太阳物质将和珠穆朗玛峰一样重。如果要把地球做成黑洞，就得把它挤压成一个半径不足2厘米的小球。多年来，物理学家一直在怀疑是不是真有这样极端的物质形态，很多人认为黑洞不过是疲惫的理论家们幻想的东西。

然而，近10年来，黑洞存在的实验证据越来越多，越来越令人信服。当然，因为洞是黑的，不可能直接用望远镜在天空搜寻。实际上，天文学家不是在找黑洞，而是找可能在黑洞事件视界外的正常发光恒星的反常行为。例如，当黑洞附近的普通恒星的外层尘埃和气体落向事件视界时，它们将加速到近光速。在这样的速度下，盘旋下落物质的内部摩擦将产生大量的热，令混合的尘埃云“发光”，向外发出可见光和X射线。因为这些辐射在视界外面，所以它们可以逃离黑洞，穿过空间，来到我们的实验室和望远镜。广义相对论预言了这些X射线所具有的各种性质。这些预言的性质的发现，为黑洞的存在提供了强有力的——尽管不是那么直接的——证据。例如，许多证据表明，在我们银河系的中心有一个巨大的黑洞，它的质量是太阳的250万倍。然而这样一个庞然大物也算不得什么，天文学家相信，在遍布宇宙的类星体（一种亮度惊人的遥远天体）的中心，可能藏着比太阳质量大10亿倍的黑洞。

史瓦西在发现他那个解几个月后，就在俄国前线染上一种皮肤病死了，那年才42岁。虽然他与爱因斯坦的引力论没打多久的交道，就令人悲伤地匆匆离去，却掀开了大自然最惊人最神秘的一层面纱。

广义相对论初露锋芒的另一个例子是关于整个宇宙的起源和演化。我们已经看到，爱因斯坦证明空间和时间有赖于质量和能量的存在。时空的扭曲影响着周围物体的运动。反过来，物体运动的具体方式通过它的质量和能量又进一步影响着时间的弯曲，而这弯曲又影响着物体的运动……宇宙的舞蹈就这样一直跳下去。通过广义相对论方程——方程源自19世纪大数学家黎曼（Georg Bernhard Riemann，

关于他我们以后再讲）在几十年前发现的弯曲空间的几何——爱因斯坦成功地定量描写了空间、时间和物质的相互演化。令他惊奇的是，当方程超越宇宙间的孤立系统（如恒星和围绕着它的行星、彗星），用于整个宇宙时，会得到一个惊人的结果：整个宇宙空间必然随时间变化。就是说，宇宙的结构要么在扩张，要么在收缩，但绝不会静止不变。关于这一点，广义相对论方程说得很明确。

这个结论即使对爱因斯坦来说也太沉重了。他推翻了千百年来人们基于日常经验建立起来的关于空间和时间本性的直觉信念，但他却根深蒂固地相信宇宙从来就是那样，永远也不会改变。为了这一点，爱因斯坦重新审查了他的方程，添加了一个著名的宇宙学常数，以帮他避免那个变化的预言，再回到令他满意的静态宇宙。然而，12年后，美国天文学家哈勃（Edwin Hubble）经过仔细的星系距离的测量发现，宇宙正在膨胀。据说（这在今天已经是科学史上有名的故事了），爱因斯坦那时又回到了他原先那个方程，把他一时添加的东西说成是他一生最大的错误。[17]尽管爱因斯坦原来并不愿意看到那样的结果，但他的理论确实预言了宇宙的膨胀。其实，在20世纪20年代初，比哈勃的发现早几年，俄罗斯天文学家弗里德曼（Alexander Friedmann）就用爱因斯坦最初的方程详细证明了，所有星系都将因空间结构的扩张而越来越快地彼此分离。哈勃以及后来无数的观测完全证实了广义相对论的这个惊人的结论。爱因斯坦因为解释宇宙的膨胀而获得了有史以来最伟大的一个理性胜利。

假如空间结构在扩张，星系在宇宙的长河里越流越远，那么我们可以追溯到从前，去认识宇宙的源头。如果时间倒流，空间结构会收缩，所有的星系也将越走越近。收缩的空间像一口压力锅，把星系压缩在一堆。温度大大升高了，星星一颗颗破碎了，形成滚烫的一堆等离子体（物质基本组元之一）。空间继续收缩，温度不断升高，原初等离子体的密度也一样地无限增大。假如时间从今天的宇宙（约150亿年）往回走，就我们所知，宇宙会挤压得越来越小。构成万物的物

质——不论地球上的汽车、房子、高楼大厦、高山大海，还是地球和月亮；不论土星、木星和其他行星，还是太阳和银河系的其他恒星；不论是仙女座的千亿颗星星，还是千亿个星河——都将被宇宙的大手捏成重重的一团。随着时间流向更远的过去，整个宇宙会缩得更小，仿佛一个橘子、一个柠檬、一粒豌豆或一粒沙，而且一直收缩下去。回到“开始”，宇宙似乎是一个点——我们在以后的章节会再来讨论这个点的图景——所有的物质和能量都挤在这一点，谁也想象不出它的密度有多大，温度有多高。

仿佛炸弹炸出无数弹片，大爆炸炸出了宇宙万物，我们该牢记这幅图景；不过，还有一点容易误会的地方。炸弹的爆炸，总是发生在空间的某个位置，在时间的某一时刻，而弹片在周围空间飞溅。对大爆炸来说，没有周围的空间。当我们回溯宇宙的开端，万物拥挤在一起，那是因为整个空间也在收缩。从橘子到柠檬到沙粒，宇宙越缩越小——我们说的是宇宙的整体，不是宇宙中的某些东西。当时间回到起点，空间也不存在了，只有那点原初的火球。所以，大爆炸是压缩的空间的喷发，它像浪潮那样扩张，把物质和能量带到今天。

广义相对论对吗

凭我们今天的技术水平，还没有在实验中发现背离广义相对论预言的事情。也许，未来更高精度的实验能发现点儿什么，从而最终证明广义相对论也不过是对大自然活动的一种近似的描写。不断提高实验精度来对理论进行系统的检验，当然是科学进步的一条途径，但不是惟一的途径。实际上我们已经看到了，寻找新的引力理论的动机并不是有什么实验违背了牛顿理论，而是因为牛顿理论与另一个理论——狭义相对论——发生了矛盾。牛顿理论的实验缺陷，是在另一个对立的引力理论（广义相对论）发现以后，从两个理论细微然而可测的偏差中显露出来的。因此，理论的内在矛盾在推动科学进步中，也起着与实验同等重要的作用。

半个世纪以来，物理学家还一直面临着另一个理论冲突，与狭义相对论和牛顿理论的冲突一样激烈。那就是，广义相对论与另一个经过极严格检验的理论，量子力学，在根本上似乎是不相容的。虽然我们讲了那么多，但这个矛盾使物理学家还不知道在大爆炸的那一刻，当空间、时间和物质统统挤成一点时，究竟发生了什么；在黑洞的中心，又究竟发生了什么。而从更一般意义说，这个矛盾在警告我们，我们关于自然的概念还存在着根本性的缺陷。一些伟大的理论物理学家曾努力过，但矛盾还没解决；它当然地成了现代物理学的中心问题。为认识这个矛盾，还需要懂一点儿量子理论的基本特征，我们接下来就去看看。

注释

1.说得更准确一点，爱因斯坦发现，只要观测局限在足够小的空间里——只要“车厢”足够小，等效原理总是成立的。原因如下：引力场的强度（和方向）会随位置发生变化，而我们想象车厢是作为一个单位在加速，所以加速度生成的是一个均匀的引力场。不过，如果车厢更小，引力场就更没有变化的空间，等效原理也就更加适用。在专业上，从加速度观点生成的均匀引力场与物质集合产生的非均匀的“真”引力场之间的差别，就是有名的“潮汐”引力场（因为它说明了月亮对地球潮汐的引力作用）。于是，本注释可以概括地说，如果观测空间很小，则潮汐引力场不会发生作用，这样加速运动和“真”引力场也就没有分别了。

2.关于所谓“刚性转盘”（即转环的更科学叫法）的分析，很容易引起混乱。实际上，在这个例子中，许多方面到今天也没达成一致意见。正文遵从了爱因斯坦本人分析的精神，现在我们还是照那个精神来澄清几点可能会令人迷惑的性质。第一点，也许有人奇怪，为什么转环的周长不跟尺子一样产生洛伦兹收缩，那样斯里姆测量的周长应该和我们原先看到的一样。不过应该记住，那环在我们的整个讨论中都是旋转着的；我们·从·来·没·有分析过它静止的情形。这样，从我们

静止观察者的立场看，我们的测量与斯里姆的测量的惟一区别是，他的尺子发生洛伦兹收缩了。我们测量时，环在旋转；我们看斯里姆测量时，环仍然在旋转。由于我们看他的尺子收缩了，所以认为他需要多测几步才能测完一个周长，那当然就比我们测量的长。只有当我们比较环在旋转和静止的性质时，环周长的洛伦兹收缩才有相对意义，但我们并不需要做这种比较。

第二点，虽然我们不需要分析静止的转环，你可能还是想知道，假如它慢慢停下来，会发生什么事情呢？看来，这时候我们应该考虑由于不同旋转的洛伦兹收缩引起的随速度的改变而改变的周长。但这如何与不变的半径相一致呢？这个问题很微妙；回答那个问题的关键一点是，世界上并没有完全的刚体。物体可以伸长或收缩，从而能够协调我们看到的伸长和收缩。假如不是这样，就会像爱因斯坦说的那样，通过熔铁在旋转运动中冷却形成的转盘将因后来旋转速度的改变而断裂。关于刚性转盘历史的详情，请看Stachel，“Einstein and the Rigidly Rotating Disk”（爱因斯坦与刚性转盘）。

3.专业的读者会发现，在转环的例子中，即在匀速旋转的参照系中，我们关注的三维弯曲空间截面可以嵌入没有弯曲的四维空间。

4.即使这样，现有的原子钟还是足以精确地测量这么小的甚至更小的时间弯曲。例如，1976年Havard-Smithsonian天文台的Robert Vessot和Mar-tin Levine与国家航空航天局（NASA）的合作者们让从弗吉尼亚沃罗普斯（Wal-lops）岛放射的侦察D火箭带了一个每小时误差不超过万亿分之一秒的原子钟。他们希望证明火箭升空（从而减弱了地球的引力作用）以后，地球上相同的原子钟（仍经历着完全的地球引力）将走得相对慢一点。研究者们可以通过微波信号的往返来比较两个原子钟的节律。他们发现，在火箭9656千米的最大高度上，原子钟比地球上的快了大约十亿分之四，符合理论的预言，精度超过了万分之一。

5.19世纪中叶，法国科学家勒维叶（Urbain Jean Joseph Le Verrier）发现，水星有一点偏离牛顿引力定律所预言的轨道。那以后的半个多世纪里，为了解释这所谓的多余的“轨道近日点进动”（通俗地说，水星在每绕太阳一圈后，不能完全回到牛顿理论预言的地方），物理学家们什么影响都想过了——例如，一颗未知行星或行星环的引力影响，一颗没有发现的卫星的影响，行星际尘埃的影响，还有太阳扁圆形（扁率）的影响——但是，没有哪个解释能赢得普遍的赞同。1915年，爱因斯坦用他新发现的广义相对论方程计算了水星的近日点，发现了（用他自己话说）令他心神荡漾的结果：广义相对论的回答与观测事实完全一致。这一成功当然使爱因斯坦对他的理论充满了信心，不过几乎所有的人都在等着证实他预言的东西，而并不满足于他解释了一个早就知道的反常现象。更详细的情况见Abraham Pais, Subtle is the Lord（New York：Oxford University Press，1982），p.253。

## 第4章　奇异的微观世界

穿过太阳系，回到地球，乔治和格蕾茜累极了。他们来到一家量子酒吧，想好好轻松一下，走出太空的影子。乔治要了他们常喝的饮料——他自己喝加冰块儿的木瓜汁，格蕾茜喜欢伏特加汽水。然后，乔治点燃一支过滤嘴雪茄，仰靠在椅子上，双手抱着头。他正要好好吸一口，才惊奇地发现雪茄不见了，从他的牙缝里溜走了。大概是从嘴里滑下来了，他这么想，然后坐起身，看它有没有在衬衫或裤子上烧个洞。他什么也没看见，雪茄不在前头。格蕾茜很奇怪，不知他在找什么。她看见那雪茄正在乔治椅子背后的柜台上。“怪了，”乔治说，“它怎么会跑到那儿去？从我脑袋里穿过去的？——可我的舌头一点儿没烧着，头上也没长什么洞啊！”格蕾茜给他检查了一下，舌头没有问题，脑袋也很正常。这时候，饮料来了。两人耸耸肩，一生经过的怪事儿够多了，今天又遇着一件。可是，酒吧里的怪事情还多着呢。

乔治看着他的木瓜汁，冰块在不停地翻滚，像碰碰车似的在杯子里撞来撞去。这回，不是他一个人在碰着稀罕了。格蕾茜端着只有乔治一半大的杯子，也看见冰块在里头碰撞，而且更加疯狂。他们分不清单独的一块冰，只见它们混乱地撞在一堆。不过，这还算不得什么；更奇怪的事情在后头。乔治和格蕾茜瞪大了眼睛看着她的杯子，看到一块冰穿透杯子落在地上。他们抓过杯子，没有发现什么不对的地方。看来，冰块没有打破杯子就穿过去了。“这一定是太空旅行后的错觉。”乔治说。尽管冰块像疯了似的，他们还是一口气把饮料喝完，回家休息了。他们匆匆离开酒吧时，竟没发现他们走过的是一道画在墙上的假门。而老顾客们已经习惯了穿墙进出，所以也没在意乔治和格蕾茜突然消失了。

百年前，正当康拉德和弗洛伊德为人类心灵带来光明时，德国物理学家普朗克（Max Planck）也向量子力学投来第一缕阳光。根据量子力学的基本概念，乔治和格蕾茜的酒吧经历，从微观的尺度看，并不需要归结到什么神秘的力量。这些陌生而奇异的事情，正是我们的宇宙在极端微小的尺度上实际发生着的。

量子框架

量子力学是我们认识宇宙微观性质的概念框架。在高速运动或大质量的情况下，狭义相对论和广义相对论曾极大改变了我们的世界观；同样，当我们考察原子或亚原子的世界时，量子力学将揭示也许更惊人的微观特性。1965年，量子力学最伟大实践者之一的费曼（Richard Feynman）写道：

有个时候报上说只有12个人懂得相对论，我不信真有这种事情。不过也许有那么一个时候，只有1个人懂那个理论，因为在文章发表以前他是惟一掌握它的人；可读过他的文章后，就会有许多人以这样那样的方式懂得相对论了，当然不止12个人。但在另一方面，我想我可以满有把握地说，没人懂得量子力学。[18]

费曼的观点是在30多年前讲的，但在今天仍然有意义。他的意思是，尽管狭义和广义相对论极大改变了我们从前认识世界的方式，但是，假如我们完全接受理论的基本原理，那么关于空间和时间的那些陌生稀奇的东西就是自然的逻辑结果。如果你能多花些工夫来思考我们在前两章对爱因斯坦理论的描述，你将（哪怕只是那么一会儿）发现我们做的那些结论都是必然的。量子力学就不同了。1928年左右，量子力学的许多数学公式和法则就已经确立了，而且从那时起，它就做出了科学史上最精确和成功的数字预言。但是，从真正意义说，运用量子力学的人不过是跟着理论的“先人们”立下的法则和公式按部就班地去计算。并不真的懂得为什么能那么做，那么做意味着什么。与相对论不同，几乎没人能与量子力学“心灵相通”。

这些说明了什么呢？是不是宇宙在微观层次的活动方式太模糊、太离奇了？难道人类世世代代从寻常尺度的现象中发展起来的思想不能完全把握“到底发生了什么”吗？或者，也许物理学家不过是因为历史的巧合建立了量子力学那么笨拙的形式，尽管计算结果是对的，却令实在的本性更加模糊了？谁也不知道。也许在将来的某一天，某个聪明人能找到一种新体系，可以完全回答量子力学里的一切“什么”和“为什么”。当然，这也是说不定的。我们能肯定的只有一件事情，那就是，量子力学绝对地不容争辩地向我们证明了，我们熟悉的寻常世界的许多基本概念，在微观领域不再有任何意义。结果，我们必须极大地改变我们的语言和逻辑，以认识和说明原子和亚原子尺度的宇宙。

在接下来的几节，我们将建立这种新语言的基础，讲述它所带来的一些惊人的奇迹。如果你跟着我们的思路看到的还是一个古怪甚至可笑的量子力学，请你记住两件事情。第一，量子力学除了在数学上是和谐的，我们相信它的惟一理由是它做出的许多预言都得到了异常精确的证实。如果有人能说出一大堆你小时候的小秘密，你还能不相信他是你走失多年的兄弟吗？第二，许多人对量子力学都有你那样的

感觉，历史上一些最伟大的物理学家也多少抱着这样的观点。爱因斯坦是完全拒绝量子力学的，甚至玻尔（Niels　Bohr）这样一位量子论的核心人物和最有力的倡导者也曾经说过，谁如果在思考量子力学时不曾有过迷惑，他就没有真正懂得它。

炉子里太热了

通往量子力学的路是从一个恼人的问题开始的。我们设想一个例子。假设你家里的电烤炉是完全隔热的，你把温度定在200摄氏度，让它有足够长的时间加热。虽然在通电以前你把炉里的空气都抽干净了，但在加热炉壁的时候它的内部还是会产生辐射波，那是与电磁波形式的光和热相同的波，既可以来自太阳的表面，也可能来自烧红的铁棒。

问题来了。我们知道，电磁波带着能量——例如，地球上的生命就全靠电磁波从太阳带到地球上来的太阳能而生存。在20世纪的开端，物理学家计算了在特定温度下烤炉内所有电磁辐射所携带的能量。根据既定的计算程序，他们遇到一个荒唐的结果：对任何温度来说，炉内的总能量都是无穷大！

人人都知道这是没有意义的——火热的烤炉可能藏着巨大的能量，但肯定不会是无限大。为了理解普朗克的解决方法，我们把这个问题再说得详细一点。当麦克斯韦电磁理论用于烤炉的辐射时，我们会发现炉壁产生的波在相对的两壁间必然是整数个波峰和波谷，如图4.1。物理学家用三个参数来描写波：波长、频率和振幅。波长是相邻两个波峰（或波谷）间的距离，如图4.2。因为两壁是固定的，如果挤在壁间的波越多，波长就越短。频率是波在一秒钟内完成的振荡循环的次数。显然，频率与波长是相互决定的：波长越长，频率越小；波长越短，频率越大。为什么呢？你可以想想摇动一端固定的长绳子。轻轻上下摇摆绳子的另一端，就能摇出大波长的波。波的频率等于手臂在一秒钟内摇动的次数，那当然是很小的。但是，如果想生

成短波，你就得发狂似地摇动绳子，就是说，你要快快地摇，结果波的频率也高了。最后，物理学家用振幅来描写波的最大深度，见图4.2。

你也许觉得电磁波抽象了一点儿，那么我们来看另一个波的例子：拨动琴弦产生的波。不同的频率对应着不同的音调，频率越高，音调也越高。琴弦的振幅取决于我们用多大力量去拨弄它，拨得越重，为波动注入的能量就越大，而大能量带来大振幅。这是可以听到的，大振幅的声音更响亮。同样，低能量对应着小振幅、小音量。

利用19世纪的热力学，物理学家可以计算炉壁向每种可能波长的电磁波注入了多少能量——或者说，炉壁是以多大的力量“激起”每一列波的。他们得到的结果很简单：每一列可能的波——不论波长多大——都带着相同的能量（一个完全由烤炉温度决定的量）。换句话讲，炉内所有可能的波动模式都是平等的，都被赋以相同的能量。



图4.1　麦克斯韦理论告诉我们，烤炉内的辐射有整数个峰谷，即整数个完整的振荡循环。

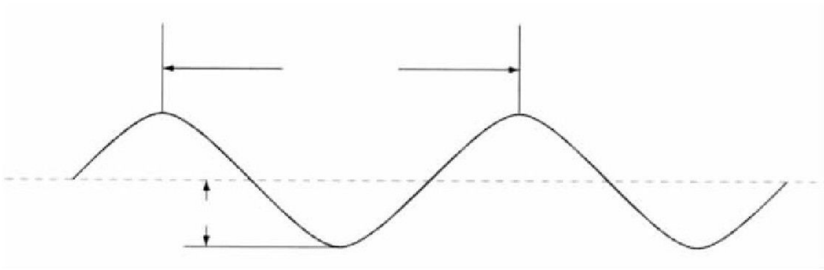


图4.2　波长是相邻两个波峰（或波谷）间的距离；振幅是波的最大高度（或深度）。

乍听起来，这是一个有趣然而却很平常的结果。实际不是这样的。它宣布已成为经典物理学的那些东西崩溃了。理由是这样的：虽然根据要求，炉内的波峰和波谷都是整数，不会有更多其他可能的波动模式，但可能的波也还是无限多的——一个波总可以有更多的波峰和波谷。因为每种模式的波带着相同的能量，所以无限多的波具有无限大的能量。这样，在世纪之交，理论物理学的油膏上飞来了一只巨大的苍蝇。19

世纪之交的能量包

1900年，普朗克提出一个激动人心的猜想，消除了无限能量的烦恼，也为他赢得了1918年的诺贝尔物理学奖。[1]为理解普朗克的猜想，我们还是来看一个例子。假设你和一大群人——无限多的人——拥挤在一个吝啬鬼老板经营的一间寒冷的大仓库里。墙上有装着令人向往的数字自动调温器，可以控制仓库里的温度。但老板收的暖气费太吓人了：如果温度调到50华氏度，每人付50元；55度，每人付55元，依次类推。你想，屋子里有无限多的人，只要打开调温器，老板就会赚得无穷多的钱。

不过，仔细读过老板的收费办法后，你发现了一个漏洞。因为老板很忙，不想找零钱（当然更不想给无限多的员工一个个地找），所

以他凭一种“信誉”方式来收费。如果谁刚好能拿出那么多钱，那他就付那么多；否则，他只需要付他尽可能拿得出来而又不需要找零的钱。你想最好人人都交费，但又不能交得太多，于是把大伙的钱都集中起来，然后照下面的方式分配：一个人全拿1分的硬币，一个人全拿5分的硬币，一个人全拿1角的硬币，一个人全拿2角5分的硬币……然后，1元、5元、10元、20元、50元、100元、1000元甚至更大（也更难得一见的）面额的钞票也照这样分别叫人拿着。你大胆把调温器开到80华氏度，然后等着老板来。老板来了，拿1分硬币的人数给他8000枚，拿5分硬币的人数给他1600枚，拿1角硬币的人数给他800枚，拿2角5分硬币的人数给他320枚，拿1元钞票的人给他80张，拿5元钞票的人给他16张，拿10元钞票的人给他8张，拿20元钞票的人给他4张，而拿50元钞票的人给他1张（因为两张就超过80元，需要找钱了）。别的人都只有1种面额的钞票——最小的也超过了应该缴纳的，所以他们不能向老板缴费。这样，老板没能拿到他期望的无限多的钱，离开时只拿走了690元。

普朗克用非常类似的办法把炉子里荒唐的无限能量减小到一个有限的大小。他是这样做的：他大胆猜想，炉子里电磁波携带的能量跟钱一样，是一小团一小团的，它可以是某个基本“能量元”的1倍、2倍、3倍……正如我们不可能有1/3分或者3/5分的硬币，普朗克声称，对能量来说，分数也是不允许的。我们的钞票是国家银行定的。那么“能量元”呢？普朗克为了寻找一个更基本的解释，建议波的“能量元”——波所能携带的最小能量——是由频率决定的。具体说，他假定一列波所具有的最小能量正比于波的频率：高频（短波）意味着大能量，低频（长波）意味着小能量。大体上讲，长波长的辐射与短波长的辐射相比，本来就缺乏动力，就像海上汹涌的浪涛都是急剧起伏的短波，只有平静的海面才会出现悠悠荡漾的长波。

关键的一点在于，普朗克的计算证明，这些允许的波的能量团消除了前面那些荒谬的无穷大的能量。不难看清这是为什么。当烤炉被

加热到一定温度时，根据19世纪的热力学理论，计算预言了每列波应该贡献的共同能量。但是，假如某些波所能携带的最小能量超过了它应该贡献的能量，它就不会对总能量有贡献——这就像那些欠老板暖气费的伙计们，因为他们手里的钞票太大了，拿不出他们该缴纳的钱。据普朗克的猜想，波的最小能量正比于波的频率。所以当我们考察炉子里的高频率（短波长）的波时，总能找到最小能量大于我们期望的能量贡献的波，它们就像那些手拿大钞票的伙计，不会为19世纪物理学要求的能量带来贡献。所以，只有有限的波能对烤炉里的总能量有所贡献，从而总能量是有限的——只有有限的伙计能缴纳他们的暖气费，老板只能收到有限的钱。钱也好，能量也好，我们靠它们不断增大的基本单位——如越来越大面额的钞票或者越来越高的频率——让无限大的结果回到了有限。[2]

消除了毫无意义的无限大结果，普朗克迈出了重大的一步。不过人们真相信他的猜想是对的，还是因为新方法计算的有限的烤炉内的能量与实验测量的结果惊人地一致。更特别的是，普朗克在计算里调节了一个参数，从而准确预言了在任意温度下测量的烤炉的能量。这个新进入计算的参数是波的最小能量单元与频率间的比例因子——也就是著名的普朗克常数，记为——大约是平常单位的千亿亿亿分之一。[20]普朗克常数这样小，说明基本h能量包的尺度也是非常小的。这也就是为什么我们觉得可以让琴弦的能量连续地变化，从而听到连续变化的琴声。虽然，照普朗克的观点，波的能量实际上是一点点传播的，但那些“点”确实太小了，一点跟着一点，仿佛是光滑连续流动的。根据普朗克的论断，这些能量包随波频率增大（即波长减小）而增大，这是解决无限大能量疑难最关键的一点。

h我们将看到，普朗克的量子假说远不仅是让我们回答了烤炉的能量问题，它还推翻了好多我们认为理所当然的世界观。因为太小，所以只有在微观世界里才会发生偏离我们日常生活的事情。但是，如果h“碰巧”大得多，那么量子酒吧[21]里的稀奇事情就会在我们身边随处

发生了。我们将看到，那些怪事儿在微观世界里确实是发生了。

能量包是什么

普朗克引进他那革命性的能量包并没有什么根据。不论他自己还是别的人，除了知道它能用以外，找不到一点儿令人信服的理由说明它为什么是对的。物理学家盖莫夫（George Gamow）曾经说过，大自然似乎喜欢喝酒，一喝就是一瓶，要么一滴也不喝，绝不会点点滴滴到天明。[22]1905年，爱因斯坦找到了一个解释，因为这个发现，他获得了1921年的诺贝尔物理学奖。

爱因斯坦的觉悟来自他对所谓“光电效应”的思考。1887年，德国物理学家赫兹（Heinrich Hertz）第一次发现，当电磁辐射（光）照在某些金属上时，金属会发出电子。这件事情本身并不特别值得注意。金属的一个特性就是，它的某些电子只是松散地束缚在原子里（这也是为什么金属是良好的导电体）。光照在金属表面时，会将能量释放出来，就像在阳光下我们会觉得皮肤暖洋洋的。这些能量会激发起金属里的电子，一些松动的电子就可能完全脱离金属表面跑出来。

但是，当我们更仔细地来研究射出电子的性质时，光电效应的奇异特征就表现出来了。乍看起来，你可能以为如果光的强度增大了（光更亮了），射出的电子的速度就会增大，因为入射电磁波的能量大了。但事实不是这样的。虽然这时候射出电子的数目增大了，但它们的速度并没有改变。另一方面，实验却发现，在入射光的频率增大时，射出电子的速度确实会增大；同样的，如果光的频率降低了，电子的速度也会减小。（对电磁波谱的可见部分来说，频率的增大相当于光从红色变到橙色、黄色、绿色、蓝色、青色，最后到紫色。频率比紫色光更高的是看不见的紫外线，然后是X射线；频率比红光还低的光是看不见的红外辐射。）实际上，如果入射光的频率减小了，会出现射出电子为零的情形。这时不论光源多么强大炫目，电子都只停

留在金属表面。由于某种未知的原因，入射光的颜色——而不是总能量——决定着电子是否发射出来，并且决定着射出电子的能量。

为明白爱因斯坦是如何解释这个难题的，我们还是回到那家大仓库。这时，仓库里的温度是80华氏度，暖洋洋的。我们想象那老板还有个毛病，他不喜欢孩子，让15岁以下的儿童都住在阴暗的地下室，大人们可以从仓库四周的回廊看到他们。另外，如果孩子们想走出仓库，只有一个办法，就是向门卫缴纳8角5分的出门费。（那老板可真不是东西！）大人们只能从回廊上向孩子们扔钱，而他们还是像以前缴暖气费那样分别拿着不同的钞票。现在来看会发生什么事情。

拿一分硬币的人先开始往下一点儿一点儿地扔。但是太小了，离孩子们需要的出门费还远着呢。因为孩子太多，都争着来抢扔下去的钱，即使拿硬币的大人扔得再多，也不够让哪一个孩子凑足他需要的85分。拿5分、1角、2角5分硬币的人也会遇着同样的麻烦。硬币不管扔了多少，孩子能抢到一个就算幸运了（大多数是空欢喜一场），谁还能拿够离开所需要的85分呢？不过，拿钞票的人接着开始往下扔了，虽然是一元一元地扔，总数也不多，但只要孩子能幸运地拿到一张，他马上就可以走了。如果那人把钞票放在桶里，一桶一桶往下放，那么每次就能有更多的孩子离开，而且每个孩子把钱交给门卫后还能剩15分。不管扔的钞票有多少，结果都是这样的。

光电效应里发生的事情差不多也是如此。根据前面讲的那些实验事实，爱因斯坦建议用普朗克的波动能量包来重新描绘光的图景。在他看来，一束光其实可以认为是一股光粒子流——后来化学家刘易斯（Gilbert Lewis）为光的微粒起了一个好听的名字，光子（在第2章光子钟的例子里，我们已经用过这个概念了）。为了有一个量的感觉，我们拿灯泡为例。根据光的粒子观，一只普通100瓦的灯泡每秒钟大概会发出1万亿亿（$10^{20}$）个光子。爱因斯坦用这个新概念提出了光电效应背后的微观机制：他指出，当一个电子被足够能量的一个光子击中时，它就会从金属表面逃逸出来。那么，是什么决定单个光

子的能量呢？爱因斯坦跟着普朗克的引导，提出每个光子的能量正比于光波的频率（比例因子是普朗克常数）。

像那些孩子离开地下室需要起码的出门费，光子必须具备一定的能量才可能将电子从金属表面解放出来。（孩子们会争抢扔来的钱，而一个电子也几乎不可能同时被几个光子击中——多数电子根本碰不上光子。）假如入射光的频率太低了，每个光子就没有足够力量激活电子。就像一枚枚硬币，扔得再多也救不了孩子，低频的光束（从而低能量的一个个光子）不论多强，也解放不了一个电子。

只要钞票来了，孩子们就可以离开仓库；同样，只要照在金属表面的光有足够的能量“元”，电子就可以脱离出来。正如拿钞票的人通过每次多扔下一些（如装在桶里）来增大总钱数，一定频率的光束也可以通过增加光子数来增大总的强度。钞票越多，能离开的孩子越多；同样，光子越多，脱离金属表面的电子也越多。不过请注意，从金属表面逃逸出来的电子的能量余额仅取决于击中它的光子的能量——而这能量是由光束的频率（而不是总强度）决定的。每个解放的电子带着相同的能量——也就是具有相同的速度——不论照射的光有多强。这就像离开仓库的孩子，不论大人扔下来多少钞票，每个孩子都还剩下15分。更多的钱只不过可以让更多的孩子离开；更多的总能量也不过是多解放一些电子。如果想让离开的孩子多带些钱，让逃逸的电子跑得更快，我们必须增大钞票的面额，提高入射光的频率——就是说，增大照在金属表面的光子所具有的能量“元”。

这些都与实验事实完全一致。光的频率（颜色）决定着射出电子的速度；光的总强度决定着射出电子的数量。这样，爱因斯坦证明了，普朗克的能量包猜想实际上反映了电磁波的一个基本特性：电磁波由粒子即光子组成，是一束光的量子。因为波是由这样的基元构成的，所以能量也就以“元”为单位了。

爱因斯坦的发现是一大进步，但我们接下来会看到，事情并不像

表面那样简单。

是光还是粒子

大家都知道，水——当然还有水波——是由大量水分子组成的，那么，光波由大量粒子（即光子）组成还有什么奇怪的吗？是的，但奇怪的是别的东西。我们知道，牛顿在300多年前就讲过光是粒子流组成的，所以这想法一点儿也不新鲜。但是，牛顿的一些同行，特别如荷兰物理学家惠更斯（Christian Huygens），却不赞同他的观点，他们认为光是波。争论了许多年，最后，英国物理学家托马斯·杨（Thomas Young）在19世纪初做的实验证明牛顿错了。

杨做的著名的双缝实验大致可以用图4.3来说明。费曼常说，量子力学的一切都可以从这个简单实验的思索中得到，所以我们应该好好来讨论它。从图4.3可以看到，光照在开了两条缝的一块薄薄的障碍物上。照相板用来记录通过缝隙的光——照片上的区域越明亮，说明通过的光越多。实验让光分别通过单缝和双缝，然后比较它们在照相板上留下的图像。

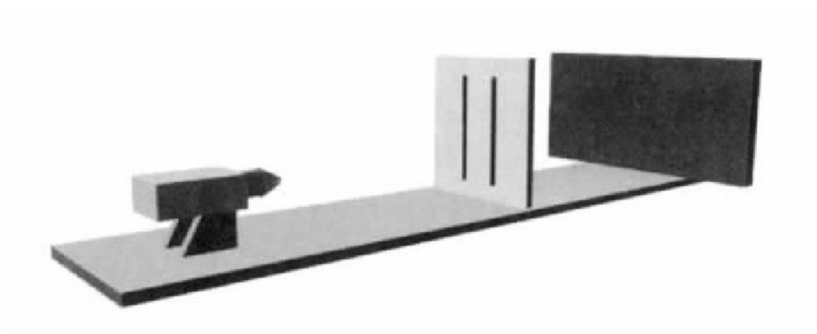


图4.3　在双缝实验中，光照在开了两条缝的薄障碍物上，通过单缝和双缝的光都将记录在后面的一块照相板上。

假如关闭左缝，打开右缝，相片将是图4.4的样子。这很好解释。因为打在照相板上的光一定是穿过右缝的，所以集中照在相片的

右边。同样，如果关闭右缝，打开左缝，相片会像图4.5的样子。假如两条缝都打开，那么照牛顿描绘的光的粒子图像，照相板应该像图4.6的样子，是前两个图的综合。大致说来，如果我们把牛顿的光微粒看成打在墙上的一颗颗小弹丸，那些通过缝隙的光就会集中在与缝平行的两个小区域。另一方面，光的波动图景在双缝打开时会表现出极不相同的景象。我们来看看吧。

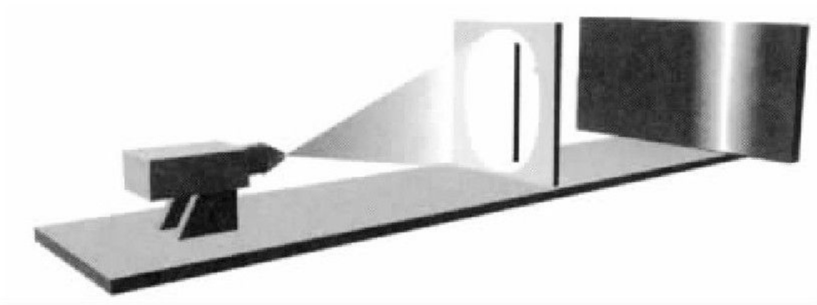


图4.4　右缝打开的实验，结果显示在照相板上。

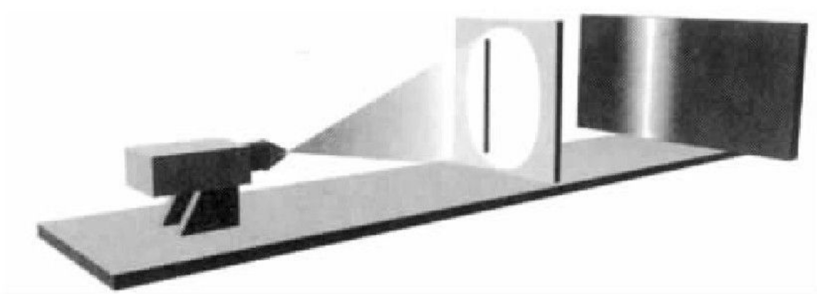


图4.5　与图4.4类似，不过开的是左缝。

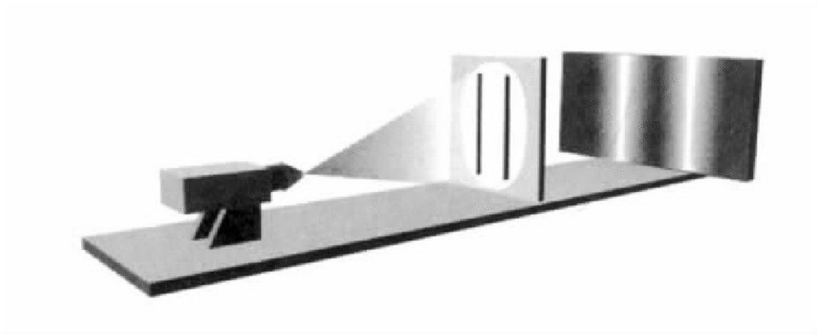


图4.6　牛顿的光的粒子观点预言，当双缝都打开时，图像应该是图4.4和图4.5的综合。

我们先考虑水波，光波的结果也是一样的，不过水更容易想象。当水波涌向障碍，穿过缝的波会一圈圈向外展开，就像一颗石子儿扔进了池塘，如图4.7。（拿一张开着两缝的卡片放在一盆水里，很容易做这个实验。）当波从两缝展开，相互重叠时，会发生一件有趣的事情。假如两个波峰相遇，那一处的水波会增高；它是两个峰的高度之和。假如两个波谷相遇，水凹陷的深度也同样会增加。最后，假如来自一缝的波峰与来自另一缝的波谷相遇，它们会彼此抵消。（实际上，消声器就利用这个道理——它测出输入的声波波形，然后生成与之“相对”的波，使它们相互抵消，消除不需要的噪音。）除了这些极端重叠的波——如波峰遇波峰、波谷遇波谷和波峰遇波谷——在它们之间还有大量部分增强或减弱的波。假如你跟许多伙伴分别坐在小船上，平行于障碍物排成一列，那么在水波经过时，你们会感觉到不同程度的颠簸。在波峰（或波谷）相遇的地方，颠簸会很强烈；在波峰与波谷相遇的地方，颠簸会很微弱，甚至一点儿也没有，因为那里的波被抵消了。

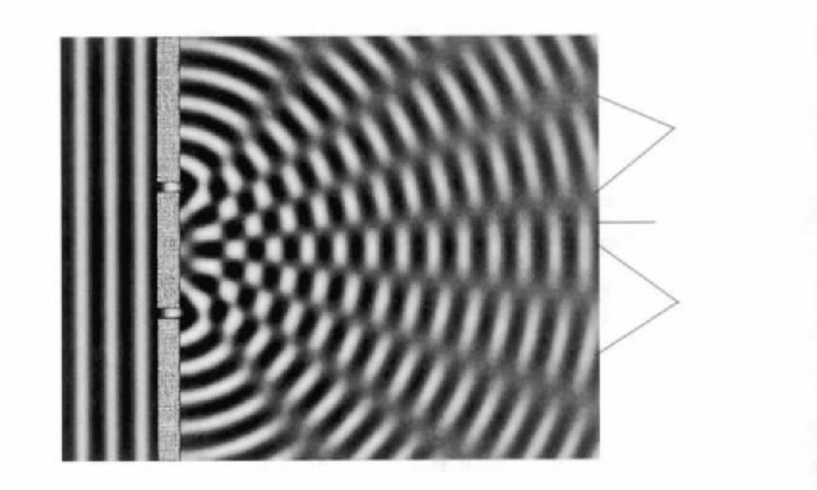


图4.7　源自两缝的水波一圈圈向外扩展，有些地方波动加强，有些地方波动减弱。

因为照相板记录了光打在板上的强弱情况，将上面关于水波的论证用于光的波动图像，我们可以发现当两条缝都打开时，照相板应该是图4.8的样子。图中最明亮的区域是来自两缝的波峰（或波谷）相遇的地方，暗区域则是一个波的波峰与另一个波的波谷相遇的地方，那里的波相互抵消了。这种光的明带与暗带交错的序列，就是有名的波的干涉模式。图4.8的照片与图4.6的大不相同，因此我们可以用这个具体的实验来判别光的粒子图像和波动图像。杨做了类似的实验，结果与图4.8的相符合，从而证实了波动图像。牛顿的微粒说失败了（不过，多年以后，物理学家才接受了这一点）。后来，麦克斯韦为流行的光的波动观奠定了坚实的数学基础。

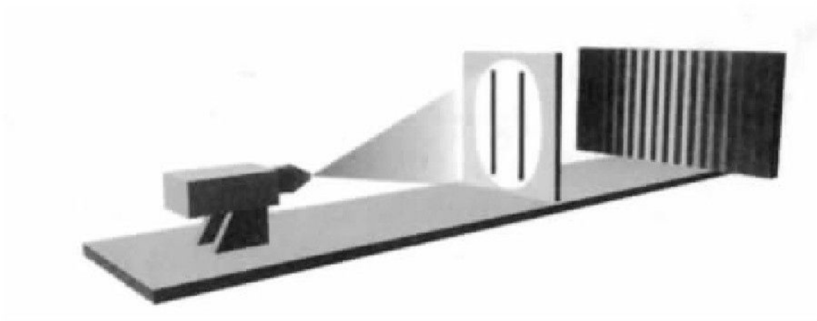


图4.8　假如光是一列波，那么当双缝都打开时，来自每一条缝的波将发生干涉。

但是，打倒了牛顿的神圣引力理论的爱因斯坦，现在却似乎又通过他的光子复苏了牛顿的光的粒子模型。当然，我们也面临着跟他同样的问题：粒子的图像如何能够解释图4.8中的干涉模式呢？也许，你马上会想到一种解释：水不是由$H_2O$分子——水“粒子”组成的吗？但大量分子形成的分子流却能生成水波，表现出图4.7所示的干涉特性。那么，我们似乎有理由猜测，光的波动特性（如干涉模式）也可以来自光的粒子图像，只要光的粒子（光子）数足够大。

可微观世界更加难以捉摸。实际上，即使图4.8里的光源强度越来越弱，直到光子以缓慢的速率（例如每10秒钟发出一个）逐个地打在障碍上，在照相板上仍然会产生像图4.8那样的结果：只要等待足够长的时间，大量分离的光子将穿过缝隙，每一个光子都记录在它打在照相板的那一点，这些点将形成图4.8里的干涉图样。这是很令人惊讶的，相继通过缝隙然后独立打在照相板上的一个个光子，如何能够“协同地”打在照相板上产生干涉波的明暗条带呢？照传统的思维，我们以为一个光子要么通过左缝，要么通过右缝，从而应该看到图4.6的样子，但事实却不是这样的。

假如这些事实没能让你感到惊讶，那可能因为你以前就知道了，或者你对什么都漠不关心，要么就是上面讲的还不够生动。为了避免后面这种情况，我们在这里以稍微不同的形式再把它讲讲。我们先将左缝关闭，让光子一个个地通过障碍，有的通过了，有的没有通过。

那些落到照相板上的一个个光子的确会生成图4.4那样的图形。然后，换一块新的照相板重做一次实验，不过这回让两条缝都打开。当然，你可能认为这只不过让更多的光子通过障碍打到照相板，因而板面会出现比第一次实验更多的光亮区域。但是，实验过后拿照片来看时，你会发现，除了意料中的有些原来暗的地方现在明了，还有些原来明的地方现在却暗了，如图4.8。我们增大了落到照相板的光子的数目，却同时减少了某些区域的亮度。看来，在时间上分离的一个个光子能以某种方式相互抵消。就是说，原来能通过右缝打在图4.8的照相板上某个暗带的光子，在左缝打开时却过不去了（因为这一点那些带才成了暗的）。想想看，这事有多奇怪。一小束光是否通过一条缝，完全取决于另一条缝是否打开，世界上怎么可能有这样的事情？这真像费曼讲的那样奇怪：假如朝照相屏幕打枪，当两条独立的缝开着时那些独立射出的子弹似乎彼此吞没了，结果失去了好多目标——而那些目标在一条缝开着时都是被打中了的。

这些实验说明爱因斯坦的光粒子大不同于牛顿的光微粒。尽管光子也是粒子，它们却不知怎么也表现着光的波动特征。粒子的能量由波的特征——频率——决定，从这个事实我们开始隐约感到波与粒子的奇特统一。而光电效应和双缝实验则真的把那统一实现了。光电效应证明光具有粒子特性，双缝实验证明光能表现波的干涉特性。两个事实结合起来证明了光同时具有粒子和波的特性。在微观世界里，我们必须摆脱在宏观世界形成的直觉——事物要么是粒子，要么是波；我们应该接受这样的事实，它可能既是波，也是粒子。我们现在该明白费曼说的那句话了，“没人懂得量子力学”。我们可以讲什么“波粒二象性”，可以将它表达为数学形式，以令人惊讶的精度来描写现实的实验。但是，我们真的很难从更深的直觉的水平去认识微观世界的奇异特征。

## 物质都是波

在20世纪开头的二三十年里，许多大理论物理学家都在不遗余力

地建立一个数学上合理而又有物理学意义的体系，来认识那些还隐藏着的事物的微观特性。例如，在哥本哈根的尼耳斯·玻尔的领导下，从炽热的氢原子发出的光的性质得到了很好的解释。但是，这同其他20年代中叶以前的工作一样，不过是19世纪的旧观念与新量子概念的临时凑合，而不是一个和谐的关于物理宇宙的认识体系。与牛顿定律或麦克斯韦电磁理论那清晰的逻辑体系相比，这些部分发展起来的量子理论还处在混沌状态。

1923年，年轻的法国贵族德布洛意（Louis de Broglie）为这场量子论战增添了新内容，他因此赢得1929年的诺贝尔物理学奖。在爱因斯坦狭义相对论的逻辑链的启发下，德布洛意提出，波粒二象性不仅适用于光，也适用于物质。他的论证大致是这样的：爱因斯坦的 $E=mc^2$联结了质量与能量，普朗克和爱因斯坦又把能量与波的频率联系起来，那么这两点的结合意味着物质也该具有波一样的形态。沿着这条思路仔细走下去，德布洛意提出，既然量子理论说明波动的光可以用粒子来描写，那么我们通常以为是粒子的电子，也应该可以同样有效地用波来描写。爱因斯坦很快就喜欢了德布洛意的思想，认为那是他的相对论和光子思想的自然结果。即使如此，它还是离不开实验的检验。实验很快就来了，是戴维逊（Clinton Davisson）和革末（Lester Germer）做的。

20世纪20年代中叶，贝尔电话公司的两个实验物理学家戴维逊和革末研究电子束如何从镍反射回来。与我们有关的细节是，镍晶体在实验中起着图4.8中双缝的作用。——实际上，完全可以认为那就是一个双缝实验，只不过以电子束代替了光束。我们下面就从双缝实验的观点来讨论。戴维逊和革末让电子通过双缝打在磷光屏上，屏幕闪烁一个亮点记下电子的位置——这也就是在电视机内部所发生的——他们发现了令人惊奇的事情：出现了像图4.8那样的干涉图样。于是，他们的实验证明了电子也表现出干涉现象，那正是波的标志。屏幕上的黑点是电子不知怎么在那里“消失了”，就像水波的波峰与波谷

在那儿相遇。即使把电子束减弱，例如，弱到10秒钟射出一个，那一个个电子仍然会形成明暗相间的条带。每一个电子都像光子那样，以某种方式相互“干涉”——说干涉，是因为它们在一定时间内重新形成了与波相联系的干涉图样。我们不可避免地会得到这样的结论：电子除了我们熟悉的粒子形态而外，还被赋予了波的特征。

尽管我们只谈了电子，类似的实验证明一切物质都具有波的特征。但是，为什么我们现实经历的事物却是硬邦邦的而一点儿都不像波呢？这问题好，德布洛意写下了一个物质波波长的公式，它表明波长正比于普朗克常数h。（更准确地说，波长等于h除以物体的动量。）因为h很小，所以计算出的波长与寻常尺度相比也小得可怜。这也是为什么只有在微观的考察中才能直接看见物质的波动性。巨大的光速c遮挡了时间和空间的本性，微小的h则令寻常世界的物质隐藏了它们波动的一面。

什么波

戴维逊和革末发现了电子的干涉现象，使电子的波动特征实在地显露了出来。但那是什么波呢？奥地利物理学家薛定谔（Er-win Schrödinger）先提出那些波是“抹开了的”电子，这有点儿电子波的意思，但太模糊了。当我们将某样东西抹开时，它总会东一点西一点的；然而，对电子来说，谁也没见过半个、1/3个或者其他分数的电子。这样，我们很难把握抹开的电子到底是什么。1926年，德国物理学家玻恩（Max Born）提出了另一种建议，极大地修正了薛定谔的电子波解释，他的新解释——经过玻尔和他的同事们的发扬光大——在今天仍然伴随着我们。玻恩的解释是量子理论最奇异的特征之一，不过已经得到了大量实验数据的证实。他指出，应该从几率的观点来看电子波。波的振幅（更准确些说，是振幅的平方）大的地方，电子出现在那儿的可能性越大；波的振幅小的地方，电子出现在那儿的可能性越小。图4.9是一个例子。

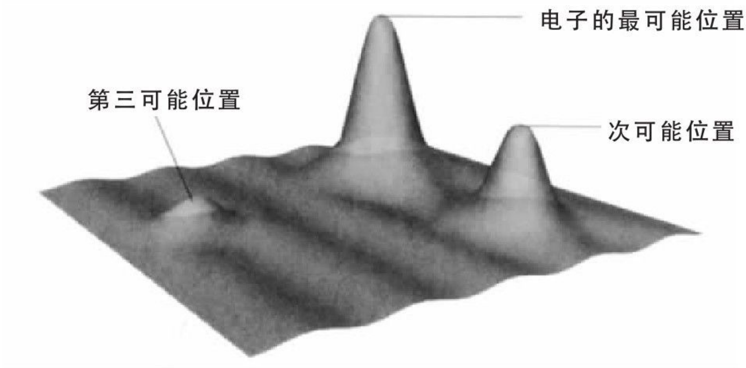



图4.9　哪里最容易找到电子，哪里的电子波最大；而在波小的地方发现电子的机会也越小。

这真是奇特的想法。基础物理学的建立与几率何干呢？我们只是在赛马、扔硬币和轮盘赌的场合才习惯几率，那不过反映了我们知道得不完备。例如，在轮盘赌中，假如我们完全知道轮盘的速度，小球的硬度和质量，它落在盘中的位置和速度，以及小格子的具体构成材料等等，假如我们有足够强大的计算机来完成需要的计算，那么，根据经典物理学，我们能够预知那小球会滚落在什么地方。赌牌的趣味正在于人们不可能在押赌注以前判断所有的信息和进行必要的计算。但是，我们在轮盘赌遇到的几率问题一点儿也没有反映宇宙活动的任何特别基本的东西。而量子力学却把几率概念植入了宇宙的深处。根据玻恩的理论和后来半个多世纪的实验，物质的波动性质意味着我们应该从根本上以概率的方式来描述物质本身。对宏观物体来说，如咖啡杯或轮盘，德布洛意的公式说明它们的波动性质微不足道，所以在大多数普通情形，我们可以完全忽略与物体相关的量子力学几率。但是，在微观水平上，我们知道我们最多不过能说在特定地方找到一个电子的几率。

几率解释的好处在于，当电子波像其他波那样活动时——例如，穿过障碍物，形成各种波荡漾开去——并不是说电子本身裂成了碎片，而是说电子可能以较大的几率在许多位置出现。实际上，这就是说，假如我们以绝对相同的方法重复某个有电子参与的实验，我们不

会总是得到相同的结果。例如，我们可能发现一个电子在不同的位置出现。而且，不断重复实验将得到各种不同的结果，在某一位置发现电子的次数，将由电子几率波的形状决定。假如A处的几率波（准确地说，应该是几率波的平方）是B处的2倍，那么经过一系列实验以后，在A处找到的电子应该是B处的2倍。我们不可能预言精确的实验结果，我们最多只能预言某个结果可能出现的几率。

即使如此，只要能从数学上决定几率波的精确形式，我们就能通过多次重复某个实验来观测某一结果发生的可能性，从而检验那些几率的预言。德布洛意的建议提出没几个月，薛定谔就迈出了决定性的一步。他写下了一个方程，能决定几率波的形状和演化——我们今天把那几率波叫波函数。薛定谔的方程和几率波的解释很快就做出了令人惊讶的精确的预言，于是，到1927年时，经典物理学的纯真时代结束了，宇宙不再是一只精确的大钟。过去我们总以为，宇宙间的一切事物都照一定的节律运动，它将从过去某个时刻走向它惟一注定了的终结。根据量子力学的观点，宇宙也遵照严格准确的数学形式演化，不过那形式所决定的只是未来发生的几率——而不是说未来一定会发生什么。

许多人感到这个结论太困惑，甚至完全不能接受。爱因斯坦就是这样的一个人。他在警告量子力学的拥戴者时，说过一句在物理学史上鼎鼎有名的话："上帝不会跟宇宙玩儿骰子。"他觉得，几率在基础物理学中出现是因为某种说不清的理由，像在轮盘赌中那样，几率出现是因为我们的认识从根本上说还不够完备。在爱因斯坦看来，宇宙没有给靠机会实现的未来留下空间。物理学应该预言宇宙如何演化，而不仅是预言某个演化发生的可能性。但是，越来越多的实验——有些令人不得不相信的实验是在爱因斯坦去世以后做的——证明，爱因斯坦错了。英国物理学家霍金（Stephen Hawking）曾说过，在这一点上，"爱因斯坦糊涂了，而量子理论是对的。"[23]

不过，关于什么是量子力学的争论今天仍在继续着。每个人都知

道怎么用量子理论的方程来做精确的预言，但是，关于几率的意义，关于粒子如何“选择”它的未来，还没有一致的认识；甚至，我们还不知道粒子是不是真的选择了一个未来；也许，它会像树枝那样分开，向着不断膨胀着的平行宇宙展开它各个可能的未来。这些观点本身都应该写一本书来讨论，实际上也有好多精彩的书以这样那样的方式思考量子理论的问题。但有一点是明白的，不论我们如何解释量子力学，从日常的立场看，它都不可否认地证明了，宇宙建立在一些奇异的原则上。

我们从相对论和量子力学得到的教训是，当我们深入追寻宇宙的基本行为时，我们会遇到许多意料之外的事情。我们需要大胆提出深层次的问题，更需要巨大的勇气来适应和接受它们的答案。

费曼的图景

理查德·费曼是继爱因斯坦以来最伟大的物理学家之一，他完全接受了量子力学的概率论核心，但在第二次世界大战后的几年里，他提出了一种强有力的新方法来思考这个理论。从数值预言的角度看，费曼描绘的图景与以前的理论完全一致，但他的公式迥然不同。我们把它放到电子双缝实验中来讨论。

图4.8的困惑在于，我们原以为一个电子要么穿过左缝，要么穿过右缝，于是我们预料结果应该是图4.4和图4.5的综合，即图4.6。从一条缝穿过去的电子本不应该关心是不是还存在另一条缝；然而它却似乎真的多少受到了另一条缝的影响。结果表现的干涉模式说明，某种对两条缝都有“感觉”的东西相遇融合在一起了，即使电子一个个地经过，结果还是那样。为了解释这个现象，薛定谔、德布罗意和玻恩为每一个电子赋予一个几率波。电子的几率波像图4.7的水波一样，“看见”了两条缝，从而也像水波那样干涉、融合。像图4.7中波浪涌起的地方，是几率波融合增强的地方，也是电子最可能出现的地方；像图4.7中波动平缓的地方，是几率波融合相消的地方，也是电

子几乎不会出现的地方。电子一个个打在荧屏上，按照这种几率波的形态分布，从而形成图4.8的干涉图样。

费曼的方法与众不同。传统观念认为，电子要么经过左缝，要么经过右缝，费曼却向它提出了挑战。可能有人会想，那是事物活动的基本特性，向它挑战岂不是太傻了。不过，你真能看见缝与屏之间的事情从而确定电子是从哪条缝穿过来的吗？是的，你能看见，但这时你就在挑战实验了。因为，你总得做点儿什么才能看到电子——例如，用光照它，也就是让光子从它反射回来。在日常的尺度，光子从树木、图画等物体上反射回来，像感觉不到的微小探针，几乎不会给那些相对的庞然大物的运动状态带去任何影响。但是，电子是小得可怜的东西，不论你怎样小心翼翼去发现它从哪条缝过来，照在它上面的光子总会影响它以后的运动。运动的改变将改变实验的结果。假如你正好通过干扰实验决定了每个电子来自哪一条缝，实验结果便从图4.8变成图4.6！量子世界保证，一旦确定了电子从哪条缝经过，两缝间的干涉现象也就消失了。

尽管据我们平常的经历，一个电子似乎应该通过一条缝，但费曼的挑战最终还是赢了——早在20世纪20年代后期，物理学家就意识到，任何想证实量子世界的基本物理量的尝试，都会破坏实验的结果。

费曼宣布，每个到达荧屏的电子实际上穿过了两条缝，这听起来是很疯狂的，但接下来还有更狂更野的事情。费曼说，每一个电子，从源到荧屏上某一点，实际上同时经历了所有可能的路径，其中的几条画在图4.10。电子以良好的次序通过左缝，同时也以良好的次序通过右缝；它可能朝着左缝去，却突然在空中调头走向右缝；它可能前后犹豫，最后通过一条缝，它还可能远征仙女座星系，然后又回来穿过一条缝到达荧屏。总之，它就这样什么路都可能经历——在费曼看来，电子要同时去“发现”连结起点和终点的每一条可能的路径。

费曼证明，他能为每一条路径赋予一个数，这些数的联合平均将给出与波函数计算相同的几率结果。这样，在费曼的图景中，不需要为电子联系一个几率波。实际上，我们得想象另一种同样奇怪的东西。电子——通常被看做一个完完全全的粒子——到达荧屏某一点的几率由到达那一点的所有可能路径的联合效应来决定，这就是费曼著名的量子力学的“路径求和”方法了。[3]

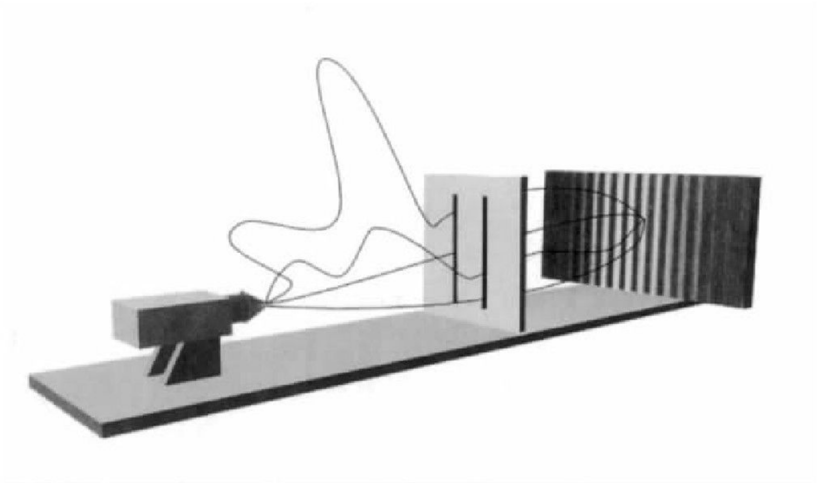


图4.10　根据费曼的量子力学形式，应该认为粒子在起点到终点间的每一条可能路径上运动。图中画出了一个电子从源到荧屏的几条可能的路径。注意这个电子实际上通过了两条缝。

在这一点上，我们学过的经典东西无能为力了：一个电子怎么能够同时经历不同的路径——而且还是无限多个呢？这似乎是很有力的反对理由，然而量子力学——关于我们世界的物理学——却要求我们把这寻常的抱怨抛到脑后。用费曼方法计算的结果符合波函数的方法，也都符合实验。我们应该允许大自然自己决定什么是合理的，什么是不合理的。正像费曼讲的，“[量子力学]描写的自然从常识看是荒唐的，但它完全符合实验。所以，我希望你们也能够那样接受她——自然本来就是荒唐的。”[24]

可是，不论微观尺度下的世界多么荒唐，在寻常的尺度下，事物

还应该回到我们所熟悉的普遍状态。为此，费曼证明，当我们考察大物体的运动时——如棒球、飞机、行星或者其他一切比亚原子粒子大的东西——他为每条路径赋值的法则保证，所有路径在求和时会彼此抵消，最后只留下惟一的一条路径。就是说，在考虑大物体时，无限多的路径里只有一条是有意义的。那也就是在牛顿运动定律中出现的轨道。这也就是为什么我们在日常生活中看到的物体（如抛向天空的球）只沿着一条预言的惟一可能的轨道从起点走到终点。但对微观物体来说，费曼为路径赋值的法则说明，许多不同的路径常常都能影响物体的运动。例如，在双缝实验里，这些路径通过不同的缝，产生我们看到的干涉图像。所以，在微观领域里，我们不能判断电子通过了哪一条缝。干涉模式和费曼的量子力学形式特别证实了另一个事实：电子从两条缝通过了。

　　我们知道，对一本书或一部电影的不同解释或多或少有助于我们理解作品的不同方面；同样，我们也需要用不同的方法来看量子力学。虽然从预言结果看，波函数方法和费曼的路径求和方法是完全一样的，但关于事件的发生，它们是不同的思维路线。我们以后会看到，在某些问题上，这些方法都能提供不可估量的解释框架。

　　古怪的量子

　　现在你多少应该感觉到量子力学里的宇宙行为是多么奇异，假如玻尔令人眩晕的理论还没有令你着魔，那么我们现在要讲的量子至少会让你头痛一会儿。

　　我们很难直觉地把握量子力学——很难像一个在微观世界里出生长大的小生命那样看量子力学，这一点比相对论更困难。不过，理论中有一点可以作为我们直觉的导引，这个特征能从根本上将量子理论与经典理论区别开来。那就是德国物理学家海森堡（Werner Heisenberg）在1927年发现的不确定性原理。

这个原理是从你可能早就想到过的一个意见产生出来的。我们讲过，为了确定电子从哪条缝通过（电子的位置），必然会干扰电子以后的运动（它的速度）。可是，你可能会想，例如，为确定我们身边是不是有人，我们可以轻轻抚摸一下，或者亲热地拍拍他的背；那么，为什么不能用“更轻柔”的光，通过更微弱的干扰来确定电子的位置呢？从19世纪物理学的观点看，我们是能够那么做的。用更暗淡的灯光（以及更灵敏的光探测仪），我们可以不断减轻对电子运动的影响。但量子力学自身却暴露了这个论证的缺陷。当我们减弱光源强度时，也就在减少它发出的光子数。当光子一个个发射出来时，我们就不可能再把它减弱了，除非把光源关闭。这是量子力学对我们所能做到的“轻柔”所规定的基本极限。于是，在我们测量电子的位置时，总会引起哪怕是极小的速度干扰。

好了，那基本上是正确的。普朗克定律告诉我们，单个光子的能量正比于它的频率（反比于它的波长）。通过频率越来越低（波长越来越长）的光，我们能产生越来越轻柔的一个个光子。但问题来了：当波从物体上反射回来时，我们收到的信息只能在相当于波长的一个误差区域内决定物体的位置。为了直观感受这个重要事实，我们想象来确定一块巨大的浸没在水中的岩石。岩石能影响经过它的海浪。波有次序地一个跟着一个涌向岩石，从岩石旁边经过时，一个个波都将被破坏——也告诉我们岩石就在那里。波的一个个起伏是一列波的最基本单位，就像直尺上的精细刻度，所以通过考察一个个波是如何被搅乱的，我们就可以在一个波动周期的误差范围内，即一个波长的范围内，确定岩石的位置。在光的情形，组成它的光子大概也可以说是一个个波动周期（波动的高度由光子决定），这样，我们用光子也只能在一个波长的精度下决定一个物体的位置。

因此，我们面临着量子力学的一种均衡行为。如果用高频率（短波长）的光，我们能以更大的精度确定电子的位置。但高频率的光子能量很大，会强烈干扰电子的速度。如果用低频率（长波长）的光，

我们可以将它对电子的影响减到最小，因为光子的能量相对说来是很小的；但是，我们却将牺牲电子位置的精度。海森堡量化了这一对抗行为，在位置和速度的测量精度间建立了一个数学关系。他发现——与我们讨论的例子一致——两个精度互成反比：位置测量的高精度必然带来速度测量的大误差，反之亦然。更重要的是，尽管我们的讨论限于以一种特别的方法来决定电子的位置，但海森堡证明，不论运用什么测量仪器，采取什么测量步骤，位置与速度的测量精度间的均衡关系总是成立的。在牛顿甚至爱因斯坦的体系中，我们都是通过位置和速度来描写粒子的运动；但量子力学不同了，它证明在微观水平上，我们不可能同时完全精确地知道那些性质。而且，一个量知道得越精确，另一个量就越不精确。虽然我们这里谈的是电子，但这个思想也能直接用于大自然的一切组成因素。

为尽可能缩小量子理论与经典物理学的偏离，爱因斯坦提出，尽管量子论显然限制了人们关于位置和速度的知识，但电子仍然像我们通常认为的那样，具有确定的位置和速度。但是，在最近几十年，以已故物理学家贝尔（John Bell）为先驱的理论发现和阿斯佩克特（Alain Aspect）及其合作者们的实验结果，令人信服地证明爱因斯坦错了。电子——以及其他所有的事物——不可能同时在某个位置并具有某个速度。量子力学证明，那样的说法不仅永远得不到实验证实——我们上面解释过了——而且也跟最近得到的其他实验结果矛盾。

事实上，假如你想捕捉固体大盒子里的一个电子，为了确定它的位置，你把盒子慢慢向里挤压，这时你会发现电子变得越来越疯狂，仿佛患了幽闭症，在盒子四壁间撞来撞去，速度越来越大，越来越难预料。不过，大自然是不会让它的“骨肉”走向死路的。我们曾想象，在量子酒吧里，h比它在真实世界里的值大得多，从而寻常的事物也能遭遇量子效应。例如，乔治和格蕾茜酒杯里的冰块也像染了幽闭症，狂野地撞得杯子砰砰作响。虽然量子酒吧是幻想的王国——实际的h小得可怜——但这种量子幽闭现象却是微观世界普遍存在的特

征。当我们考察微观粒子并将它们限定在越来越小的空间区域时，它们会变得越来越疯狂。

从不确定性原理还生出一种令人惊奇的效应：量子隧道。假如你向着3米厚的混凝土墙射出一个塑料小球，经典物理学的结论与你本能的感觉是一样的：球会反弹回来。原因是，小球没有足够的能量穿透这堵难以逾越的障碍。但是量子力学确凿地证明，在基本粒子的水平上，组成小球的粒子的波函数——即几率波——总有一小部分透过了墙。这意味着小球有小小的——但不是零——机会能穿透那堵墙，出现在墙的另一边。怎么能这样呢？原因还是在海森堡的不确定性原理。

为明白这一点，我们想象一个一贫如洗的人，他在远方的亲戚死了，给他留下一大笔遗产。然而，他没有钱买飞机票，去不了那儿。他把困难告诉了朋友：如果朋友们能帮他解决这个难题，借钱给他买张机票，他回来时可以加倍奉还。但是朋友们也没钱。不过，他忽然记起一位在航空公司的老朋友，便向他求救。那朋友还是没钱借他，但提出一个办法：他只要在到达终点后24小时内把钱电汇过来，公司的会计系统是不会发现钱是在出发以后补的。这样，他可以去继承他的财产了。

量子力学的统计过程大概也是这样的。我们已经看到，海森堡曾证明，在位置和速度的测量精度间存在一种平衡。他还证明，同样的平衡关系也存在于测量能量和测量时间的精度之间。量子力学断言，我们不能讲一个粒子在某一时刻具有某个能量。为提高能量的测量精度，必须增大测量的时间。大体上讲，这意味着在足够短的时间尺度内，粒子的能量可能疯狂地涨落起伏。所以，就像我们可以“借钱”坐飞机（只要能尽快还钱），量子力学也允许粒子“借”能量，只要它能在海森堡不确定性原理所规定的时间内把它还回去。

量子力学的数学证明，能量障碍越大，那种奇怪的微观统计行为

实际发生的几率就越小。但是，面对一块石板的微观粒子，有时的确可能借来足够的能量在瞬间穿透它原来没有足够能量进入的区域——从经典物理学的观点看，这简直是不可能的。当我们研究的东西越来越复杂，组成的粒子越来越多，量子隧道效应仍然可能发生，不过越来越困难，因为那需要所有粒子都能幸运地穿过隧道。但是，乔治的雪茄消失在脑后，酒杯里的冰块从杯壁漏出，乔治和格蕾茜从酒吧的墙上穿过——这些奇异的事情都能发生。在我们假想的量子酒吧，h很大，那样的量子隧道到处都是。但在真实的世界里，h很小，假如你频繁走近一堵墙，根据量子力学的几率法则，你需要等待我们今天宇宙的年龄这么久，才能找到一个好机会穿墙而去。只要有耐心（还得长寿），你迟早会出现在墙的另一边。

不确定性原理抓住了量子力学的核心。我们通常认为的一些毫无疑问的基本事实——例如物体有确定的位置和速度，有确定的能量，确定的动量——现在看来不过是因为普朗克常数太小而在寻常世界表现的一些特例。更重要的是，当我们把量子思想用于时空结构时，“引力的锦绣图景”将暴露致命的缺陷，把我们引向过去百年里的第三次物理学大冲突。

注释

1.虽然普朗克的工作确实解决了无限大能量的难题，但那显然不是他研究的动机。相反，普朗克在寻求认识一个密切相关的问题：关于烤炉——准确说是“黑体”——的能量在不同波长范围如何分布的实验结果。关于这段历史的更详细情况，有兴趣的读者可以参考Thomas S.Kuhn（库恩），Black-Body Theory and the Quantum Discountinuity，1894～1912（Oxford, Eng.：Clarendon，1978）。

2.说得更准确一点，普朗克证明了，最小能量超过它们平均能量贡献（据19世纪的热力学）的波会以指数形式衰减；我们所考察的波

的波长越大，这种衰减将迅速增大。

3.应该说明的是，费曼的量子力学方法可以导出以波函数为基础的一般方法，反过来也可以；于是，这两种方法是完全等价的。不过，每种方法所强调的概念、语言和解释是迥然不同的，尽管答案绝对相同。

## 第5章　渴望新理论：广义相对论与量子力学

在过去的100年里，我们对物理宇宙的认识已经非常深入了。量子力学和广义相对论的理论工具使我们懂得了很多事情，我们能够很好地预言发生在原子、亚原子领域乃至星系和星系团尺度的物理现象，甚至我们还能认识整个宇宙本身的结构。这是了不起的成就。真的，一种普普通通的生命，困在一颗小小的行星上，在一个普普通通的星系的边缘，绕着一颗普普通通的恒星旋转，却能凭他们的头脑和实验，去发现和理解物理宇宙中某些最神秘的特征。不过，物理学家天生是难得满足的，他们还要去把最最幽深和基本的东西揭示出来，这就是霍金讲的，认识“上帝的大脑”的第一步。[25]

很多证据说明，量子力学和广义相对论没能达到最深层的认识。它们通常的适用范围是不同的。大多数情形，要么用量子力学，要么用广义相对论，不会同时需要它们。然而，在某些极端条件下，事物质量大而且尺度小——例如在黑洞的中心、在大爆炸时刻的宇宙——为了得到正确的认识，我们既需要量子力学，也需要广义相对论。可是，量子力学与广义相对论，一个像火药，一个像火，它们遇到一起便带来巨大的灾难。当这两个理论的方程混合起来时，好好的物理问题却得不出有意义的答案。那些无意义的答案主要是一些无限大的东西，如关于某过程的量子力学几率的预言，不是一个百分数，而是无限大。大于1的几率已经够荒谬了，无穷大的几率是什么呢？我们被迫承认，一定出了什么严重的问题。通过认真考察广义相对论和量子力学的基本性质，我们可以找出毛病在哪儿。

量子力学的核心

海森堡发现不确定性原理时，物理学在那儿拐了一个大弯，但并没有停歇下来。几率、波函数、干涉和量子，都带着认识实在性的崭新思路。不过，顽固的“经典”物理学家们还抱着一丝希望，盼着当一切都弄清楚以后，这些“离经叛道”的东西将树立一个离过去思路不远的理论框架。然而，不确定性原理把所有的“复辟”幻想都扫荡干净了。

不确定性原理告诉我们，当我们考察的距离越小、时间越短，宇宙会变得越疯狂。上一章讲过，当我们想确定基本粒子（如电子）的位置和速度时，会遇到这种情况：用更高频率的光照电子，我们能以更高精度测量它的位置，但那代价是我们的观测更多地干扰了电子的运动。高频率的光子具有更多的能量，所以像针一样“扎”在电子上，从而极大地改变了它的速度。在一间满是小孩儿的屋子里，我们会遇着同样的狂乱场面。你可以精确知道每个孩子的瞬间位置，但你却几乎管不了他们的活动——向哪个方向跑，跑多快——不可能同时知道基本粒子的位置和速度，意味着微观世界在本质上是混沌的。

尽管这个例子在不确定性与疯狂性之间建立了基本联系，它也只说明了问题的一部分。你大概会认为，不确定性只有在我们这些笨拙的自然观测者闯进了它们的场景才会出现。这是不对的。电子在小盒子里飞速地撞来撞去，这个例子可能会让你更明白那是怎么回事。即使实验者没有“直接”拿光子去“打击”电子，电子速度还是会剧烈地不可捉摸地从一点变到另一点。但是，这个例子也没能完全说明海森堡的发现所隐藏的微观世界那迷人的特征。即使在我们所能想象的最宁静的场合，例如空空如也的空间区域，不确定性原理告诉我们，从微观的角度看，那里也有大量的活动。距离和时间的尺度越小，那活动就越狂乱。

明白这一点的关键是量子的会计方法。我们在前一章看到，粒子

（例如电子）可以暂时“借”能量来克服难以逾越的物理障碍——就像人们常常可以借钱渡过难关。这是对的。但量子力学迫使我们将这类比向前推得更远。我们想象一个不得不靠借钱生活的人，他去求一个个朋友，每个朋友只能借他几天，他只得找更多的朋友，这家借，那家还，还了借，借了还——他费好大力气借来钱，不过是为了尽快把它还掉。像华尔街狂涨狂跌的股票价格一样，这位可怜的借钱者手里的钱也在瞬间经历着巨大的涨落。不过，当一切平息过后，他的账目说明他还跟当初一样，一点儿也没富起来。

海森堡的不确定性原理说，在微观的距离和时间间隔里，能量和动量也发生着类似的疯狂的涨落。即使在虚空的空间——例如一只空盒子——不确定性原理也会说能量和动量是不确定的：当我们从更小的时间尺度来看更小的盒子时，它们的涨落就更大。仿佛盒子里的空间也不得不“借”能量和动量，不断从宇宙把它们“借来”，接着又很快还回去。那么，在平静的空虚的空间区域里，哪些东西参与了这样的“交易”呢？什么东西都可能有，这真是难以想象的；不过，最终“流通”的还是能量（也包括动量）。$E=mc^2$告诉我们，能量可以转化为物质，而物质也能转化为能量。这样，如果能量涨落足够大，即使在虚空的空间里，它也可以在瞬间生成正反粒子对，例如电子与它的正电子伙伴。因为这些能量必须马上归还，所以粒子对会在瞬间湮灭，归还生成它们的能量。其他形式的能量和动量也发生着相同的事情——如其他粒子的生成与湮灭、电磁场疯狂的振荡、强弱相互作用场的涨落……量子力学的不确定性原理告诉我们，宇宙在微观尺度上是一个闹哄哄的、混沌的、疯狂的世界。费曼曾笑话过，“生了灭、灭了生——浪费了多少时间。”[26]由于能量的借与还在平均意义上相互抵消了，所以只要不是微观地去看，空虚的空间仍然显得宁静而太平。但是，不确定性原理说明，宏观的平均的眼光模糊了众多微观的行为。[1]我们很快会看到，融合广义相对论和量子力学的那个障碍，就是这里讲的那些疯狂的东西。

量子场论

20世纪三四十年代，理论物理学家们在不懈地寻找一种数学形式来描写微观世界的混沌行为，我们可以提几个杰出的名字，如狄拉克、泡利（Wolfgang Pauli）、施温格（Julian Schwi-nger）、戴森、朝永振一郎（Sin-Itiro Tomonaga）和费曼。他们发现，薛定谔的波动方程（第4章讲过）实际上只是微观物理学的一种近似描写——当我们不太深入微观的混沌时（不论实验的还是理论的），这近似是非常好的；但当我们想走得更近时，它就失败了。

薛定谔在他的量子力学方程里遗忘了一个重要的东西，那就是狭义相对论。实际上，他当初确实试过把狭义相对论包括进来，但后来发现那方程做出的预言不符合氢原子的实验观测。于是薛定谔继承了物理学的老传统，将问题分开来解决：不急着跨一大步把所有新发现的物理学东西都塞进一个新的理论；通常更好的办法是小步、小步地走，一步步地把研究前沿的最新发现囊括进来。薛定谔发现并建立的数学方程包含了实验发现的波粒二象性，但在那个初步认识的年代，他没有包括狭义相对论。[2]

但物理学家很快发现，狭义相对论对一个正确的量子力学框架来说是很重要的。这是因为，微观的混沌行为让我们看到了能量可以有多样表现形式——这个观点来自狭义相对论的$E=mc^2$的结果。薛定谔的方法忽略了狭义相对论，也就忽略了物质、能量和运动的重要性。

为了把狭义相对论与量子概念结合起来，物理学家们首先把力量集中在电磁力与物质的相互作用上。经过一系列激动人心的进步，他们创立了量子电动力学。这是后来相对论量子场论（或简称量子场论）的一个例子。说它是量子的，因为几率和不确定性的观点从一开始就融合在理论中了；说它是场论，因为它把量子原理融入了以前的经典力场的概念——在这里，那是麦克斯韦的电磁场。最后，我们说它是相对论的，因为狭义相对论也是从一开始就走进来了。（如果你

想对量子场有一个直观的认识，你可以很好地借助经典场的想象——例如，数不清的看不见的力线穿过空间——不过，那图像应该有两个特点：第一，量子场应该由粒子构成，就像电磁场由光子组成一样；第二，能量应该以粒子质量和运动的形式出现，它不停地在量子场之间往来波动，在空间和时间里不停地振荡。）

量子电动力学可以说是有史以来关于自然现象的最精确理论。我们可以从木下东一郎（Toichino Kinoshita）的例子来说明这种精确。木下是康奈尔大学的粒子物理学家，在过去的30年里，他一直在艰辛地用量子电动力学计算电子的某些具体性质。他的计算写满了几千页，最后还是世界上最大的计算机来完成的。他的努力是值得的：计算的结果在小数点后面十二位都得到了实验的证实。这绝对是抽象的理论计算与现实世界之间惊人的契合。通过量子电动力学，物理学家能在完备、实用、可靠的数学框架下巩固光子作为“最小的一束光”的角色，揭示它们与带电粒子（如电子）的相互作用。

量子电动力学的成功激励其他的物理学家在20世纪六七十年代去发展一门类似的新的量子力学方法，以认识弱、强和引力的作用。结果证明，对弱力和强力来说，这是一条硕果累累的道路。通过与量子电动力学类比，物理学家构造了强力与弱力的量子场论，叫量子色动力学和量子弱电理论。“量子色动力学”这个色彩绚丽的名字，在逻辑上该叫“量子强动力学”，但那不过是一个名字而已，没有别的更深的意思。[27]另一方面，“弱电”这个名字确实概括了我们在认识自然力的长路上所树立的一座里程碑。

格拉肖（Sheldon Glashow）、萨拉姆（Abdus Salam）和温伯格（Steven Weinberg）通过他们赢得诺贝尔奖的工作，证明了弱力与电磁力可以自然地用他们的量子场理论统一起来——尽管两种力在我们周围世界的表现好像是迥然不同的。毕竟，除了在亚原子的尺度内，弱力场几乎消失了，没有一点儿作用；而电磁场——可见光、无线电波、电视信号、X射线……却是我们离不开的宏观实在物。不

过，格拉肖、萨拉姆和温伯格从根本上证明，在足够高的能量和温度下——如在大爆炸的几分之一秒内——电磁场和弱力场熔化在一起，表现出不可分辨的特征，应该更准确地叫弱电场。当温度下降，电磁力与弱力便结晶似地分离开来（分离的过程实际从大爆炸时就开始了），具有与高温下不同的形式——这样一个过程就是有名的“对称破缺”，我们会在以后慢慢讲——从而在我们今天冰冷的宇宙中表现得迥然不同。

好了，现在我们知道，到20世纪70年代，物理学家已经对四种力中的三种（强、弱、电磁）做了合理而成功的量子力学描述，还证明了其中的两种（弱和电磁）实际上有着共同的起源（弱电力）。在过去的20年，物理学家做了大量实验，通过那三种力的自我表现和它们在第1章介绍的物质粒子中的作用，来检验量子力学的处理方法。理论安然面对了所有的实验挑战。实验家们测量了19个特别的参数（表1.1中的粒子质量和第1章注释1中补充的力荷、表1.2中的三个力的强度，以及几个别的我们不需要讨论的参数），理论家将这些数引进物质粒子和三种力的量子场理论，结果，这个微观宇宙的理论做出的预言与实验符合得好极了。在我们今天的技术条件下——所达到的能量可以将物质粉碎到百亿亿分之一米的大小——理论都是正确的。因为这一点，物理学家把这个关于引力外的三种力和三族物质粒子的理论叫做标准理论，或者，更多的时候称它是粒子物理学的标准模型。

信使粒子

根据标准模型，强力和弱力的场也有最小的组成粒子，就像电磁场以光子为最小组成一样。我们在第1章曾简单讲过，最小的强力单元是胶子，而最小的弱力单元是弱规范玻色子（或者，更准确地说是W和Z玻色子）。标准模型要求我们把这些力的粒子看成没有内部结构的——在这样的框架下，这些粒子全都是基本的，跟那三族物质粒子一样。

光子、胶子和弱规范玻色子提供了它们所组成的力的微观传递机制。例如，当一个带电粒子排斥另一个带同性电荷的粒子时，我们大体上可以想象每个电子都裹着一个电场——一团“电的云雾”——每个粒子感觉的力就源自那“两团云”的排斥。不过，更准确的量子图景却多少有些不同。一个电磁场由一群光子组成，两个带电粒子间的相互作用实际上是光子在两个粒子间往来“出没”的结果。两个带电粒子通过交换小小的光子而相互影响，这个过程有点儿像两个溜冰的人连续不断地相互传球，通过传球，两个人的运动状态都将受到影响。

然而溜冰者的比喻却有一个大毛病，那就是传球总是“排斥性的”——它使两个人越离越远。不同的是，带相反电荷的粒子也通过交换光子发生相互作用，而那电磁力却是“吸引的”。看来，光子似乎并不是力本身的传递者，它只不过传递“消息”，告诉粒子该如何响应那个力。对同性电荷的粒子，光子带来的消息是“离开”；而对异性电荷的粒子，那消息是“走近”。因为这一点，我们有时说光子是电磁力的信使粒子。同样，胶子和弱规范玻色子分别是强力和弱力的信使粒子。把夸克锁在质子和中子里的强力起源于一个个夸克交换胶子。可以说，胶子真就像把这些亚原子粒子紧紧粘在一起的“胶”。弱力决定着粒子的某种放射性衰变的嬗变过程，它的中介是弱规范玻色子。

规范对称性

你大概已经看到了，我们关于自然力的量子理论的讨论还遗漏了一个奇异的角色，那就是引力。物理学家靠那个方法成功地处理了其他三种力，你可能会建议他们去寻找一个引力的量子场理论——在那个理论中，引力场的最小单元引力子应该是它的信使粒子。乍看起来，这该是一个特别合时宜的建议，因为另外三种力的量子场论告诉我们，在它们与我们在第3章遇到的引力的某个方面之间存在着诱人的相似。

回想一下，引力让我们能够把所有的观察者——不论他们的运动

状态如何——看做绝对平等的。即使通常认为在加速运动的人，也能说自己是静止的，因为他可以把感觉到的力归结为他处在一个引力场中。在这个意义上，引力强调对称：它保证所有可能的观察者的观点以及所有可能的参照系都是同样有效的。同样，强力、弱力和电磁力也都通过对称性相联系，虽然那些对称比同引力相关的对称要抽象得多。

为了粗略体会这些难以捉摸的对称性原理的意思，我们考虑一个重要的例子。我们记得，在第1章后面的注释里，每个夸克都带着三种“颜色”（我们想象那是红、绿、蓝，当然这不过是一些标签，与通常看到的色彩没有一点儿关系），这些颜色决定着夸克该如何“响应”强力，就像电荷决定着对电磁力的响应一样。在得到的所有数据的基础上，我们建立了夸克间的一种对称性，那就是，同色夸克（红与红、绿与绿或蓝与蓝）间的相互作用都是相同的，不同色夸克（红与绿、绿与蓝或蓝与红）间的相互作用也是相同的。实际上，我们还从数据发现了更令人惊奇的事情。假如夸克携带的三种颜色（三个不同的强荷）都“转移”了（用假想的色彩来说，大概意思是，红、绿、蓝转移成黄、青、紫），甚至它们每时每刻从一个地方到另一个地方都在不停地转移，夸克之间的相互作用却一点儿也不会改变。因为这一点，我们说宇宙体现着一种强力对称性：物理学不因力荷的转移而改变——或者说，物理学一点儿也不知道力荷转移了。这很像我们说球体现着旋转对称性，因为不论我们在手里怎么转，不论转多大的角度，球看起来都是一样的。由于历史的原因，物理学家也说强力对称是规范对称的一个例子。[3]

我们来看关键的一点。在广义相对论中，为使所有可能的观察立场都处于对称地位，必须要求存在引力。类似地，从外尔（Hermann Weyl）20世纪20年代以及杨振宁和米尔斯（Robert Mills）20世纪50年代的工作发展起来的规范对称性也要求存在另外一些力。根据杨振宁和米尔斯的观点，那些力场能完全补偿力荷的转移，从而完全地保

证粒子间的物理相互作用不会改变。这很像一个灵敏的环境控制系统，通过彻底补偿任何外来的影响而使一个区域内的温度、气压、湿度保持为常数。对与夸克的色荷转移相关的规范对称来说，需要的力不是别的，正是强力。就是说，如果没有强力，物理学在上面说的色荷转移下会发生改变。这一发现表明，虽然引力和强力有许多不同的性质（回想一下，引力比强力弱得多，而作用范围却远得多），它们确实还有某种相同的特征：它们的存在是为了让宇宙享有特别的对称性。而且，相同的论证也适用于弱力和电磁力，它们的存在关联着另外的规范对称性——所谓的弱与电磁的规范对称。因此，四种力都直接联系着对称性原理。

四种力的这个共同特征预示着我们在这节开头的建议是很有希望的。那就是，为了把量子力学融入广义相对论，我们应该寻找一种引力的规范场理论，就像物理学家已经发现了的其他三种力的成功量子场理论一样。近些年来，这个思想激励了一大批有名的物理学家满怀热情地踏上寻找之路，但一路上困难重重，还没有谁走到尽头。我们来看那是为什么。

广义相对论与量子力学

广义相对论适用于巨大的天文学尺度。在那样的距离，爱因斯坦的理论说明，没有物质意味着空间是平直的，像图3.3画的那样。为了把广义相对论与量子力学融合起来，我们现在必须转移关注的焦点，去考察空间的微观性质。在图5.1中，我们说明了如何一点点去暴露越来越小的空间结构。开始的时候，看不出什么来；看图中底下的三层，空间结构几乎是一样的形态。从纯经典的立场看，我们以为这样平直稳定的空间图景会一直保持到任意的距离尺度。但量子力学完全改变了这种想法。万物都摆脱不了不确定性原理所规定的量子涨落——引力场也不例外。虽然经典理论认为虚空间没有引力场，但量子力学证明，引力场尽管在平均意义上等于零，实际上却因量子涨落而波荡起伏。另外，不确定性原理还告诉我们，关注的空间越小，看

到的引力场起伏越大。量子力学展现了一个没有绝望的世界，越是狭小的地方，越是浪花飞溅。

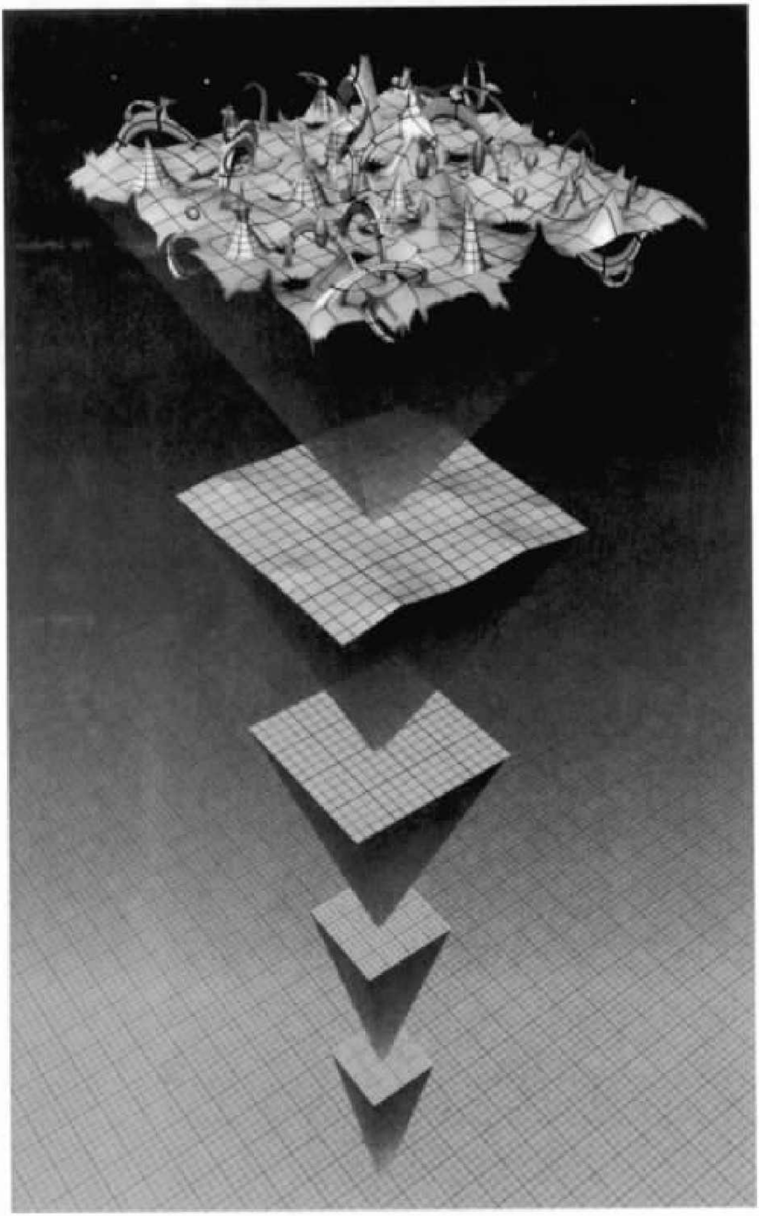


图5.1　逐级放大空间区域，显露它的超微观特性。为了融合广义相对论与量子力学，我们将在最高的地方面对汹涌的量子泡沫。

引力场通过空间的弯曲表现出来，而量子涨落通过周围空间越来越强烈的扭曲表现自己。在图5.1中我们看到那种扭曲隐约出现在第四层。向更小的距离尺度逼近，我们会在第五层遇到随机的量子力学波动，那里的引力场表现出极强烈的空间弯曲，一点儿也不像我们在第3章画过的弯曲的橡皮膜。实际上，它像图顶那样，到处是混沌的卷曲。惠勒发明了一个名词量子泡沫来描绘这种超微的空间（和时间）里表现出的混沌状态——它描绘了一幅陌生的图画，传统的一些概念，如左和右、前和后、上和下（甚至过去和未来），都失去了意义。还是在这样的小尺度上，我们才发现广义相对论与量子力学原来是不相容的。广义相对论的核心原理——光滑的空间几何的概念——被小距离尺度的量子世界的剧烈涨落破坏了。在超微尺度上，量子力学核心的不确定性原理与广义相对论核心的空间（以及时空）的光滑几何模型是针锋相对的。

实际上，那矛盾是很具体地表现出来的。把广义相对论和量子力学融合起来的所有计算，都得到一个相同的答案：无限大。这是一个当然的信号，告诉我们做错了事情，该让老师打手掌心了。[4]广义相对论的方程平息不了量子泡沫的喧嚣。

不过，应该看到，当我们回到寻常尺度（在图5.1中从上往下看），小距离尺度上剧烈的随机涨落会平息下来——像那位被迫借钱的人把钱还了，银行的账户没留下借钱的痕迹——这时，宇宙结构的光滑几何学又变得精确了。我们有过这样的经历：从远处看到的一幅色彩均匀光亮柔和的图画，走近一看，却跟光滑的画面大不相同，原来它不过是一点点色斑，每一点都是分离的。但是你得注意，只有在离图很近，一点点地看，你才会发现它原来是离散的；而从远处看时，它是光滑的。同样，除了在超微观的尺度下，时空结构都表现得很平坦，这也是为什么广义相对论在足够大的距离（和时间）尺度——与许多典型的天文学问题相关的尺度——能做得很好，而在小距

离（和时间）尺度上却产生那么多矛盾。广义相对论核心的光滑和轻微弯曲的几何图像，在大尺度上证实了；但在推向小尺度时，却被量子涨落破坏了。

广义相对论和量子力学的基本原理使我们能够在某个很小的尺度上进行计算，不过，低于那个尺度时，图5.1里的可怕现象会表现得很明显，计算不能再往前走了。因为标志量子作用强度的普朗克常数太小，描写引力本来强度的引力常数也太小，它们构成一个更小的几乎难以想象的普朗克长度：十亿亿亿亿分之一（$10^{-33}$，小数点后面32个零）厘米。[5]图5.1最高层描绘的就是在普朗克长度下的超微观的宇宙景观。为了对那尺度有一个具体的认识，我们想象，把一个原子放大到我们的宇宙尺度，那么普朗克长度也不过是一棵普通的树的高度。

于是我们看到，广义相对论与量子力学间的冲突只是发生在宇宙相当隐蔽的地方；因为这一点，你当然可以问，那些问题值得去忧虑吗？有些物理学家也很明白那个问题，但他们还是在研究需要的时候，在典型尺度远远超过普朗克长度的问题上，快乐地运用广义相对论和量子力学。而另外一些物理学家则深信，我们那两块物理学基石根本就搭配不起，并不因为超微观的尺度才暴露了问题。他们认为，这个矛盾指出了我们对物理宇宙认识的根本缺陷。这种看法源于一个不能证明然而深入人心的世界观：如果在最深最基本的水平上认识宇宙，宇宙应该能以一个各部分和谐统一、逻辑上连贯一致的理论来描述。不论那矛盾对各人的研究是不是根本性的，有一点是肯定的，那就是，大多数物理学家很难相信，我们对宇宙最深层的认识的理论基础是由两个虽然有力然而搭配不起的数学框架拼接起来的。

为了能让两个理论协调起来，物理学家做过大量的尝试，他们以这样那样的方法，要么修正广义相对论，要么修正量子力学；虽然一次次的努力通常都胆识惊人，但结果却是一个失败跟着一个失败。

终于，超弦理论来了。[6]

注释

1.你可能还在疑惑：虚空间的区域里还能发生什么事情吗？重要的是应该知道，不确定性原理为空间能有“多空”做了限制，与我们平常讲的虚空是不同的。例如，关于场的波扰动（如在电磁场中传播的电磁波），不确定性原理指出，波的振幅和振幅改变的速度也服从一个类似于位置和速度的反比关系：振幅确定得越精确，它改变的速度就越不精确。现在，我们说一个空间区域是空的，意思是没有波经过这个区域，所有的场都是零。说得啰嗦一点（但却是有用的），我们可以讲，通过这个区域的所有波的振幅都为零。但是，如果我们对振幅知道得那么精确，则不确定性原理告诉我们，振幅的改变是完全不确定的，我们可以说它有任意的数值。但如果振幅改变了，就说明它们在下一时刻不再是零，即使空间区域还是“空的”。不过，平均说来，场还是零，因为它在某些地方为正，而在其他地方为负，区域的总能量是不会改变的。当然，这只是在平均意义上说的。量子不确定性说明场的能量即使在空虚的空间区域里也是涨落的，我们关心的空间区域距离和时间尺度越小，看到的涨落就越大。因为瞬间涨落得到的能量将通过$E=mc^2$转化为瞬时的粒子和反粒子对，然后它们很快湮灭。结果，能量在平均意义上仍然没有改变。

2.虽然薛定谔写的那个原始的包括了狭义相对论的方程不能准确描写氢原子中的电子的量子力学性质，但很快发现它在其他场合还是有意义的，实际上今天我们还在用它。不过，薛定谔发表方程的时候，克莱茵（Oskar Kle-in）和戈登（Walter Gorden）已经赶到前头了，所以那个方程叫“克莱茵-戈登方程”。

3.我们为数学爱好者们多说两句。基本粒子物理学用的对称性原理一般是以群为基础的，而特别应该注意的是李（S.Lie）群。基本粒子表现为不同群的表示，而它们的时间演化方程则需要体现相关的对

称信息。对强力来说，对称性叫SU（3）（类似于普通的三维旋转群，不过是作用在复空间的），给定夸克的3种颜色在一个三维表示下变换。正文里讲的颜色转移（由红、绿、蓝到黄、青、紫）实际上就是作用于夸克“色坐标”的一个SU（3）变换。规范对称则是另一种对称性，它的变换可能依赖于时空：这时候，在不同空间位置和时刻“转动”夸克的颜色，结果也会不同。

4.在发展引力以外的那3种力的量子理论过程中，物理学家也遇到过一些无穷大结果的计算，不过，他们后来发现这些无限大的东西可以通过一种叫重整化的技术来消除。在结合引力与量子力学的工作中出现的无限大要严重得多，重整化技术也无可奈何。只是在近些年，物理学家才意识到，无限大结果的出现说明我们把理论用到了超越它应用范围的地方。我们现在的目标是寻找一个实用范围在原则上无限的理论——一个“终极的”、“最后的”理论——所以，物理学家想找一个不论分析的物理系统多么极端也不会出现无限结果的理论。

5.我们可以通过简单的论证，即通过物理学家所说的量纲分析，来认识普朗克长度。思路是这样的：如果一个理论是由一组方程建起来的，为将理论与实在相联系，抽象的符号必然与自然的物理特性相结合。特别是，我们必须引进一个单位系统，就是说，假如一个符号代表着长度，我们应该能有一个标准来说明它的数值。例如，一个方程中的长是5，我们当然需要明白它是5厘米、5千米还是5光年。在涉及广义相对论和量子力学的理论中，单位的选择是以下面的方式自然出现的。广义相对论依赖于两个自然常数：光速c和牛顿引力常数G。量子力学依赖于一个自然常数h。看看这几个数的单位（例如，c是速度，应该表达为距离除以时间，等等），我们可以发现，组合 $\sqrt{\hbar G/c^3}$ 具有长度的单位，实际上，它等于$1.616\times10^{-33}$厘米。这就是普朗克长度。它既包含了引力和时空的量（G和c），也与量子力学（h）有关，所以，在任何联合广义相对论与量子力学的理论中，它都确定了一个测量标准——一个自然的长度单位。我们在文中用“普朗克长度”通常是一个大概的意思，指长度在$10^{-33}$厘米的几个数

量级范围内。

6.目前，除了弦理论，还有人正积极地以别的方法来结合广义相对论和量子力学。一个方法是牛津大学彭罗斯（Roger Penrose）的扭量理论，另一个方法——部分是在彭罗斯的激发下兴起的——是宾夕法尼亚州大学的Ab-hay Ashtekar所引导的新变量方法。虽然本书以后不再更多讨论这两个方法，但人们越来越觉得它们可能与弦理论有着深刻的联系，而且，同弦理论一起，三个理论都在为同一个结果，为结合广义相对论与量子力学，而磨刀霍霍。

# 第三篇
# 宇宙交响曲

## 第6章　万物都是音乐：超弦理论的基础

当人们考虑同宇宙有关的一些问题时，音乐总是我们选择的方向。从毕达哥拉斯古老的“天球的音乐”到“自然的和谐”，千百年来一直引导着我们去追寻天体平和运行的天然乐音和亚原子粒子混沌的喧嚣。自超弦理论发现以来，音乐的幻想成了惊人的现实，因为这个理论认为，微观世界里到处是小小的琴弦，它们不同的振动便合奏出宇宙演化的交响曲。根据超弦理论，变化的劲风吹遍了一个充满琴弦的宇宙。

另一方面，标准模型却把宇宙的基本组成看做一点一点的没有内部结构的粒子。虽然这个方法很有力量（我们说过，标准模型提出的几乎每个微观世界的预言在百亿亿分之一米的尺度上都得到了验证，那已经达到今天的技术极限了），却成不了完备的或最后的理论，因为它没有包括引力。而且，把引力囊括进它的量子力学框架的尝试也都失败了，原因是在超微尺度下——也就是在小于普朗克长度的距离下——空间结构将出现剧烈的涨落。这个尚未解决的矛盾激励着人们去寻找一个更深的自然理论。1984年，还在玛丽王后学院的格林（Michael　Green）和加州理工学院的施瓦兹（John　Schwarz）提出了第一个令人信服的证据，说明超弦理论（或简称为弦理论）可能是我们寻找的那样东西。

超弦理论革命性地修正了我们对宇宙的超微观性质的理论描述

——物理学家慢慢发现，那修正正是我们需要的，它使爱因斯坦的广义相对论与量子力学完全相容了。根据弦理论，宇宙的基本构成要素不是点粒子，而是有点儿像细橡皮筋的上下振动着的一堆丝线。不过，别让这名字给骗了：它不像一根普通的弦，本身也由分子和原子组成；弦理论的弦被认为是深藏在物质核心里的。根据理论，弦是构成原子的粒子的超微观组成元。弦理论的弦小得可怜，平均大约是普朗克长度的尺寸，所以即使用我们最灵敏的仪器来检查，它们看起来也跟像点一样。

不过，简单地用弦来代替点粒子作为万物的基元，已经产生了深远的结果。第一点，也是最重要的一点，弦理论似乎解决了广义相对论与量子力学间的矛盾。我们将看到，弦在空间延展的本性，是把两个理论结合到一个和谐框架里来的一个关键的新要素。第二点，弦理论提供了一个真正的统一理论，因为所有物质和力都来自同一个基元：振动的弦。最后一点，我们在后面几章还会更彻底地讨论，那就是，除了上面提到的成绩，弦理论又一次极大地变革了我们对时空的认识。[28]

弦理论简史

1968年，年轻的理论物理学家维尼齐亚诺（Gabriele Veneziano）正在费力弄清实验观测到的强核力作用的各种性质。他那时是欧洲核子研究中心（CERN）的研究人员，在瑞士日内瓦的欧洲加速器实验室，对那些问题已经研究了好多年。一天，他突然有了一个惊人的发现。令他惊奇的是，著名瑞典数学家欧拉（Leon-hard Euler）在200年前因纯粹数学目的构造的一个不太起眼的公式——所谓的欧拉β函数——似乎一下子就描写了强相互作用的大量性质。维尼齐亚诺的发现将强力的许多性质纳入一个强有力的数学结构，并掀起一股热浪，用欧拉β函数和它的各种推广去描写从全世界收集来的不同原子碎片的数据。不过，维尼齐亚诺的发现从某种意义上说是不完整的。欧拉的β函数似乎很有用，但没人知道为什么；就像一个学生靠记忆

用了公式，但不知道它的意义和证明。那时，β函数还是一个等待解释的公式。到1970年，情况变了。芝加哥大学的南部阳一郎（Yoichiro Nambu）、尼尔斯·玻尔研究所的尼尔森（Holger Nielsen）和斯坦福大学的苏斯金（Leonard Susskind）揭示了藏在欧拉公式背后的物理学秘密。他们证明，如果用小小的一维的振动的弦来模拟基本粒子，那么它们的核相互作用就能精确地用欧拉函数来描写。他们论证说，这些弦足够小，看起来仍然像点粒子，所以还是能够与实验观测相符。

虽然强力的弦理论直观、简单，也令人满意，但不久人们发现它也有失败的地方。20世纪70年代初，高能实验已经能探索更深层的亚原子世界。实验表明，弦模型预言的某个数直接与观测结果相矛盾。这时候，作为点粒子量子场理论的量子色动力学也在发展着，它在描写强力时获得了压倒一切的成功，弦理论当然也就黯然失色了。

大多数粒子物理学家认为，弦理论已经被扔进了科学的垃圾堆。不过，有几位虔诚的研究者还在守着它。例如，施瓦兹觉得“弦理论的数学结构太美了，还有那么多奇妙的性质，一定关系着什么更深层的东西。”[29]物理学家发现的一个弦理论问题是，它似乎“管得太多”了。这个理论中振动的弦的图像具有类似胶子的性质，这一点证实了它原来是一个强力理论的宣言。但是，除了这些，它还包含着多余的信使粒子，似乎与强力的任何实验观测都不相干。1974年，施瓦兹和巴黎高等师范学院的谢尔克（Joël Scherk）迈出了大胆的一步，使这一显然的缺陷成了优点。他们在研究了那令人疑惑的像信使粒子一样的弦振动模式后，发现它完全符合假想的引力的信使粒子——引力子。尽管这些“最小的引力单元”从来没有发现过，理论家还是能预言它们应该具有的某些性质，而施瓦兹和谢尔克则发现这些性质正好通过一定的弱振动模式实现了。在这个基础上，谢尔克和施瓦兹提出，弦理论最初的失败是因为我们不恰当地限制了它的范围。他们断言，弦理论不单是强力的理论，也是一个包含了引力的量子理论。[30]

物理学圈子里的人并没有满怀热情地欢迎他们的建议，实际上，施瓦兹说："我们的工作被普遍忽略了。"[31]在统一引力和量子力学的征途上，人们已见过太多的失败；弦理论当初在描写强力时也有过错误，在很多人看来，带着它去追求一个更宏伟的目标似乎是没有意义的。更令人失望的是，20世纪70年代末和80年代初的研究证明，弦理论和量子力学遭遇了各自微妙的矛盾。看来，引力还是"不愿意"走进宇宙的微观图景。

直到1984年，情况才有了变化。格林和施瓦兹经过10多年艰苦的遭大多数物理学家白眼、排斥的研究，终于在一篇里程碑式的文章里证明了，令弦理论困惑的那个微妙的量子矛盾是可以解决的。而且，他们还证明，那个理论有足够的能力去容纳4种基本力。这些话传遍了整个物理学世界，许许多多的粒子物理学家都停下他们的研究计划，涌向这最后一个理论的战场——为了一个古老的追求，认识宇宙最深最远的秘密。

我从1984年开始在牛津大学读研究生，虽然我为所学的量子场论、规范理论和广义相对论感到兴奋，但老同学们却普遍感觉粒子物理学前途渺茫。标准模型摆在那里，预言的实验结果那么成功，它的证实是迟早的事情，最多不过补充些细节。超越它的极限，把引力包括进来，而且要能解释它所依赖的实验事实——概括基本粒子质量的那19个数，它们的力荷，力的相对强弱，那些从实验得到却还没有理论根据的数……——一项多么可怕的使命，只有最勇敢的物理学家才敢迎接这个挑战！但是，6个月以后，气氛完全不同了。格林和施瓦兹的胜利最后也感染了一年级的研究生，身在物理学历史的伟大运动中的激情，替代了以往的忧郁。我们多数同学都攻读到深夜，就为了学会理解弦理论所需要的大量的理论物理和抽象的数学。

从1984年到1986年，是我们所谓的"第一次超弦革命"时期。在那3年里，全世界的物理学家为弦理论写了一千多篇研究文章。这些研究明确地证明，标准模型的许多特征——那是经过几十年艰难探索

发现的——简单地在弦理论的宏大结构中自然出现了。正如格林说的，“当你遇到弦理论，发现近百年来所有的重大物理学进步都能从那么简单的起点产生出来——而且是那么美妙地涌现出来——你会感觉，这个令人着迷的理论真是独一无二的。”[32]另外，我们还将讨论，对多数性质来说，弦理论的解释比标准模型更完美，更令人满意。这些成果使许多物理学家相信，弦理论正在一步步实现它的愿望，成为一个终极的统一理论。

但是，弦理论总是一次又一次地遭遇同一块巨大的绊脚石。在理论物理学研究中，我们经常遭遇的只不过是难解或难懂的方程。物理学家一般不会放弃，而是试着近似地解决它们。弦理论的情形则更加困难，连方程本身都很难确定，至今我们也只是导出了它的近似形式。于是，弦理论家们只限于寻找近似方程的近似解。经过第一次革命的巨大进步以后，物理学家发现，他们运用的近似解不足以回答挡在理论前头的许多基本问题。除了近似方法，物理学家们找不到别的具体方法。于是，有些走进弦理论的人感到沮丧，又回到他们过去的研究路线。对留下的人来说，20世纪80年代末和90年代初是他们热身的时期。弦理论像一座宝库，但锁得严严的，只能通过一个小孔看到它，可望而不可即；它那么美妙，那么有希望，在召唤着人们，但没人有打开它的钥匙。漫长平淡的日子过后总会迎来重大的发现。但每个人都明白，我们还需要强有力的新方法来超越过去的近似方法。

接下来，在南加利福尼亚召开的“弦1995年会”上，惠藤通过他那激动人心的演讲——一篇令在场的世界顶尖物理学家们大吃一惊的演讲——宣布了下一步的计划，从而也点燃“第二次超弦革命”。弦理论家们跟我们这儿讲的一样，都在费尽心力地磨炼一套新的方法，有望能克服以前遇到过的那些理论障碍。全世界的超弦理论家们的技术本领都将面临前进路上的困难的考验，而在那另一尽头的光明，虽然还很遥远，总有一天会看到的。

在这一章和接下来的几章里，我们要谈通过第一次超弦革命到第

二次超弦革命以前的研究得到的对弦理论的认识。有时我们也会用后来的眼光去看前头的东西；而最新的进展要等到第12章和第13章。

还是希腊人的原子吗

我们在本章开头讲过，图1.1也画过，弦理论宣扬的是，如果能以远远超越我们现在能力的精度去检验标准模型假设的点粒子，我们将看到，每个粒子都是单独的一根细细的振荡着的小线圈儿。

以后我们会明白，这些闭合的弦一般是普朗克长度的尺度，大约是原子核的一万亿亿分之一（小数点后面19个零）。难怪我们今天的实验还不能决定物质的微观的弦的本性：即使在亚原子粒子的尺度上看，弦也是太小太小了。我们需要用加速器来把物质能量比以前做的提高大约1000亿倍，才可能直接揭示它们是弦而不是点粒子。

我们将简单说明以弦代替点粒子会产生哪些惊人的结果。不过，还是先来讲一个更基本的问题：弦是什么做的？

问题有两个可能的答案。第一，弦是真正基本的东西——是“原子”，在古希腊人本来的意义上，也就是不可分的基元。绝对的最小的构成万物的基元的弦，像最后的那个俄罗斯洋娃娃，[33]代表着微观世界数不清的亚结构层次走到了尽头。从这点看，弦即使在空间延伸，问它们的组成也是没有意义的。如果弦是由更小的事物组成的，它们就不会是基本的。相反，如果什么东西构成了弦，它就当然可以取代弦的位置，而成为更基本的宇宙基元。用语言学的类比，我们说，段落由句子组成，句子由词语组成，词语由字母组成，那字母由什么组成呢？从语言学的立场看，字母是最基本的东西。字母就是字母，它们是书面语言基本的建筑砖块，没有更细的结构。问它们的组成也是没有意义的。同样，弦就是弦，没有比它更基本的东西，所以不能把它描写成由别的任何物质组成的东西。

以上是第一个答案。第二个答案基于目前的现实情况：我们还不知道弦理论是不是正确的大自然的最后理论。假如弦理论真的走错了方向，我们可以忘记弦和不相干的关于它们组成的问题。虽然这是可能的，但20世纪80年代中期以来的研究令人不得不相信，事情很可能不是那样的。不过另一方面，历史也确实告诉我们，每当对宇宙的认识深入一步，我们总会发现物质还有更微观的层次，还有更小的组成元素。所以，关于弦是否能成为最后的理论，还有一种可能，那就是，它们仿佛是宇宙大洋葱剥下的一层，在普朗克长度下可以看到这一层，尽管还不是最后的一层。在这种情形，弦可能由更小的结构组成。弦理论家们提出了这种可能性，也在不停地寻找这种可能性。今天，理论研究中出现了一些有趣的线索，暗示弦可能有更小的结构，但还没有确实的证据。经过艰苦的研究，总有一天我们能回答这个问题。

除了第12章和第15章的几点猜想，我们都在第一个答案的前提下讨论弦的问题——就是说，我们认为弦是大自然最基本的组成单元。

通过弦理论走向统一

标准模型除了不能把引力包括进来，还有一个缺点：不能具体解释它的那些组成。为什么大自然会选择表1.1和表1.2列的那些特别的粒子和力？为什么描写这些粒子和力的19个量具有那样的数值？你可能不禁会想，那些数和具体的性质似乎都是任意的。这些看起来很随机的组成单元的背后是否还藏着什么更深的道理？难道宇宙的这些形形色色的物理学性质真是被偶然“选中”的？

标准模型本身不可能解释这些问题，因为它把这些粒子和它们的性质当做实验观测为它输入的原始数据。在没有原始投资数据的情况下，股市的表现不能用来决定证券盈亏；同样，如果离开了那些基本粒子性质的数据输入，标准模型什么也预言不了。[1]在实验粒子物理学家一丝不苟测量那些数据以后，理论家就能用标准模型做出一些可

以检验的预言，如某些粒子经过加速器的轰击后会发生什么事情。但是，标准模型不能解释表1.1和表1.2里的基本粒子性质，就像今天的道·琼斯指数不可能解释你10年前买了多少股票。

实际上，假如实验发现了什么不同的微观世界的粒子可能与不同的力发生作用，我们只需要把不同的参数输入理论，就很容易把这些变化纳入标准模型。从这个意义说，标准模型的结构也太能“善变”了，能适应很多可能的事情，所以它解释不了基本粒子的性质。

弦理论大不一样，它是独特的牢固不变的理论大厦。它不需要输入更多的参数，只需要一个确定测量尺度标准的数（下面讲）。微观世界的一切性质都在它的解释能力之内。为明白这一点，我们先来考虑大家熟悉的弦，如小提琴的弦。每一根琴弦都可能有许多（实际上是无限多）不同的振动模式，也就是我们知道的共振，如图6.1。共振是那些峰谷正好在弦的两个端点间张开的波动模式，我们的耳朵感觉这些不同的共振，就听到不同的音调。弦理论中的弦也有类似性质，在这里，弦可能产生的共振模式是在它的空间范围内恰当展开的峰和谷。图6.2列举了几个例子。像琴弦的不同振动模式奏响不同乐音那样，一根基本弦的不同振动模式生成了不同的质量和力荷。这是最核心的一点。因为核心，我们再说一遍：依照弦理论，一个基本“粒子”的性质——它的质量和不同的力荷——是由它内部的弦产生的精确的共振模式决定的。

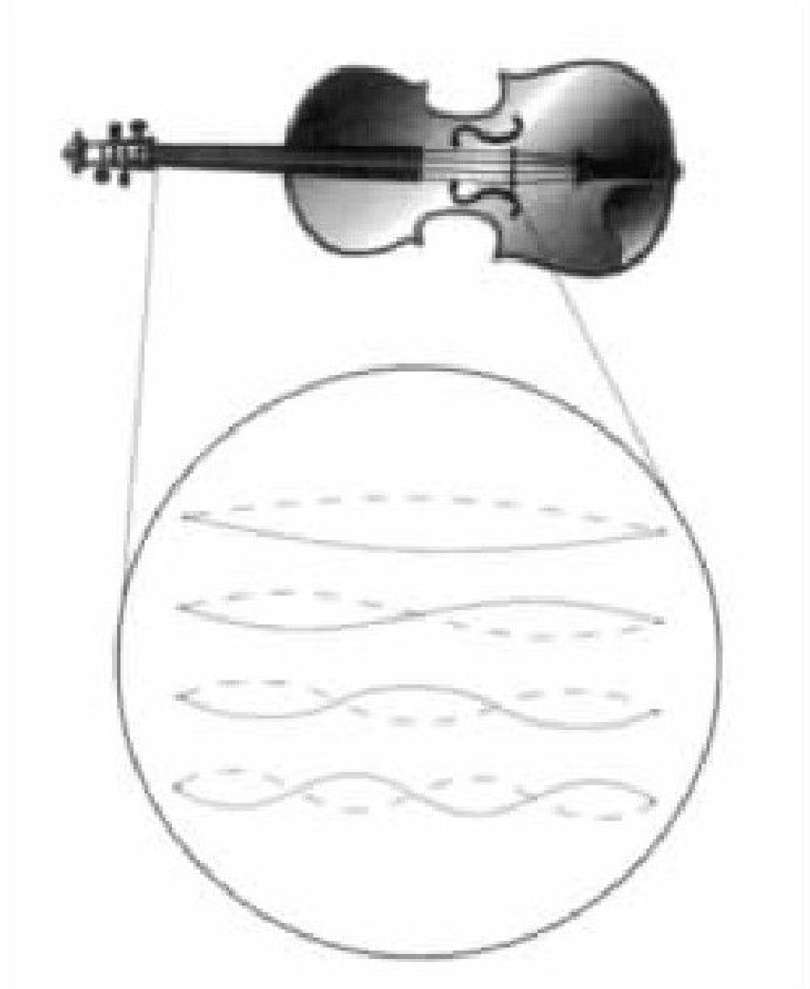


图6.1　琴弦能产生的共振模式。共振波的峰谷数目正好能满足弦的两个端点间的距离。

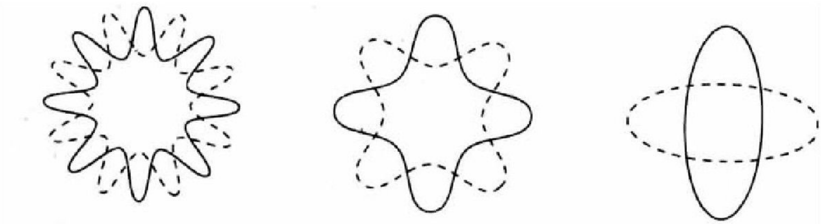


图6.2　弦理论中的闭合线圈也能以共振形式振动——类似于琴弦的共振——峰谷的数目也正好适应弦的空间长度。

弦与粒子质量的关联很容易理解。弦的某个振动模式的能量取决于它的振幅（峰谷的最大相对位移）和波长（相邻两个峰或谷之间的距离）。振幅大的和波长小的，能量较大。这与我们的直觉是一致的——振动越疯狂，那个模式的能量越大；不那么疯狂的振动，能量也会小些。我们在图6.3里列举了两个例子。这也是我们熟悉的现象。当用力拨动琴弦时，振动会很剧烈；而轻轻拨动它时，振动会很轻柔。现在来看，从狭义相对论我们知道，能量和质量像一枚硬币的两面，是同一事物的不同表现：大能量意味着大质量，大质量也就是大能量。那么，依照弦理论，基本粒子的质量决定于内在弦的振动模式的能量。质量较大的粒子所具有的弦振动较剧烈，质量小的粒子所具有的弦振动较轻柔。

因为粒子的质量决定着它的引力性质，于是我们在这里看到，弦的振动模式与粒子的引力作用之间存在着直接的关联。物理学家还发现，在弦振动模式的其他方面与其他力的性质之间，也存在着类似的关联，尽管这里涉及的论证多少要抽象一些。例如，一根弦所携带的电荷、弱荷与强荷也完全由它的振动方式决定。另外，这些关联对信使粒子本身也完全成立。如光子、弱规范玻色子和胶子等粒子，还是弦的共振模式。特别重要的一点是，在弦的振动模式中，有一种模式完全满足引力子的性质，从而保证了引力是弦理论的不可分割的一部分。[2]

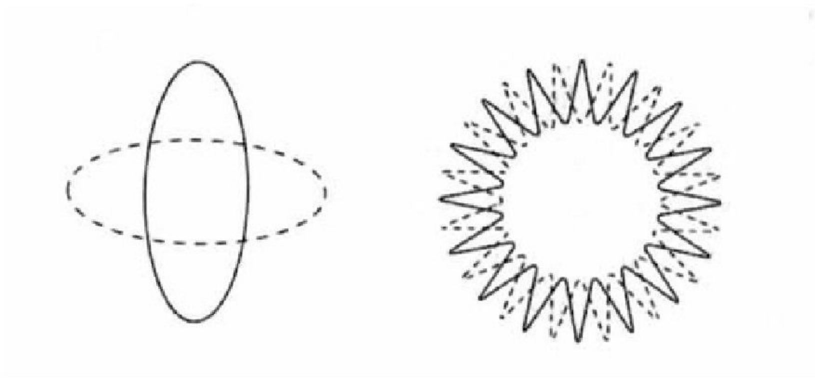


图6.3　疯狂的振动模式比轻柔的振动模式有更大的能量。

现在我们明白了，依照弦理论，每种基本粒子所表现的性质都源自它内部的弦经历着特别的共振模式。这种观点与物理学家在弦理论发现之前提出的主张是迥然不同的。照从前的观点，基本粒子间的差别大致被解释为每种粒子都是“从不同的结构里分离出来的”。虽然每个粒子都被看做是基本的，但各自被赋予的“基元”类型却是不同的。例如，电子的“基元”带负电荷，而中子的“基元”没有电荷。弦理论彻底改变了这幅图景，它宣布所有物质和力的“基元”都是相同的。每个基本粒子都由一根弦组成——就是说，一个粒子就是一根弦——而所有的弦都是绝对相同的。粒子间的区别是因为各自的弦在经历着不同的共振模式。不同的基本粒子实际上是在同一根基本弦上弹出的不同“音调”。由无数这样振动着的弦组成的宇宙，就像一支伟大的交响曲。

上面的概括说明，弦理论搭起了一个多么辉煌的真正的统一理论的框架。物质的每一个粒子，力的每一个传递者，都由一根弦组成，而弦的振动模式则是识别每个粒子的“指纹”。发生在宇宙间的每一个物理学事件，每一个过程，在最基本的水平上都能用作用在这些基本物质组成间的力来描写，所以，弦理论有希望为我们带来一个包容一切的统一的物理宇宙的描述，一个包罗万象的理论（T.O.E.）。

弦的音乐

虽然弦理论远离了以前的没有结构的基本粒子的概念，但旧的语言还很难消失，特别是在最微小的距离尺度上，过去的一些语言还为实在提供了准确的描述。所以，我们以后还是继续习惯地讲“基本粒子”，不过它的意思总是“一根根振动着的弦”。上一节我们说过，这样一些基本粒子的质量和力荷是相应的弦的振动方式的结果。这就使我们认识到，假如能够弄清基本弦的可能振动模式——或者说，“听清”它们所能奏响的“音调”——那么，我们就能解释所看到的基本粒子的性质。于是，弦理论第一次搭起了一个解释我们所观察到的自然粒子性质的框架。

这样说来，我们该“抓”一根弦来“弹”，用所有的方法去弹，以决定可能的振动模式。如果弦理论是对的，我们将发现那些可能的模式能完全产生表1.1和表1.2里的物质和力的各种粒子的观测性质。当然，弦太小了，不可能像我们讲的那样实验。不过，我们可以用数学语言在理论上弹一根弦。20世纪80年代中叶，许多弦的信奉者都认为做这些实验所要求的数学分析差不多就能解释宇宙最微观水平的每一个性质。有些热情的物理学家还宣扬，一个包罗万象的理论终于找到了。10多年过后来看，有这样信念的人也高兴得太早了。弦理论是有点儿T.O.E.的模样，但一路的坎坷还很多，我们还得不出足够精确的能与实验结果相比较的弦振动模式。所以，我们现在并不知道表1.1和表1.2总结的宇宙基本特征能不能用弦理论来解释。如我们将在第9章讨论的，在一定假设条件（我们将具体说明是什么条件）下，弦理论可以生成一个宇宙，在定性上具有与我们知道的粒子和力相符的性质，但目前还没有办法从理论导出具体的数值预言。因此，虽然弦理论的框架与点粒子标准模型不同，它能解释为什么粒子和力有我们看到的那些性质，但我们还不能把这些解释从理论中抽象出来。不过，值得注意的是，弦理论包容多、延伸远，即使确定不了具体的性质，我们还是能够认识许多从这理论生出的新物理学现象，这一点我们会

在后面的章节里看到。

在接下来的几章，我们要比较详细地讨论弦理论目前遇到的困难。还是先来大概了解一下它们。我们周围的“弦”都有着不同的张力，例如，鞋带通常比小提琴的琴弦松，而这两样又都远不如钢琴的金属弦那么有力。弦理论为了确立它的总体大小，需要的一个量就是弦圈的张力。如何决定张力呢？是这样的。如果我们拨动一根弦，那么我们将知道它的强度是怎样的，这样，我们就能像测量普通弦的张力那样来测量基本弦的张力。但基本弦太小，这种办法实行不了，还需要有更间接的办法。1974年，谢尔克和施瓦兹提出，某个特别的弦振动模式是引力子，他们找到了那种间接的方法，从而预言了弦理论的这些弦的张力。他们的计算表明，通过那假想的弦振动的引力子传递的力的强度反比于弦的张力。我们曾设想引力子传递的是引力——一种天生很微弱的力——于是他们发现，那意味着引力子的弦有巨大的张力，千万亿亿亿亿（$10^{39}$）吨，这是所谓的普朗克张力。这样看来，基本弦与我们熟悉的那些例子相比，是极端强硬的，这引出三点重要结果。

硬弦的结果

第一点，固定琴弦的两端，弦长度也就固定了；但对基本弦来说，没有琴架子来把弦固定起来。实际上，弦理论的闭弦会因为强大的张力而收缩成很微小的弦圈，详细计算表明，在普朗克张力的作用下，一根典型的弦只有普朗克长度的大小，$10^{-33}$厘米——我们以前讲过的。[3]

第二点，因为弦理论中振动圈的张力巨大，它的能量一般也是极高的。为明白这一点，我们可以想想，弦的张力越大，就越难让它振动。例如，拨动小提琴的弦很容易，拨动钢琴就要难一点儿。所以，张力不同的两根弦，虽然振动方式完全一样，也不会有相同的能量。张力大的弦比张力小的弦有更高的能量，因为赋予它更多的能量，它

才能产生运动。

这提醒我们，振动弦的能量由两样东西决定：振动的准确模式（振动越疯狂，能量越高）和弦的张力（张力越大，能量越高）。乍看起来，这可能令人想到，如果我们让振动越来越轻柔——振幅越来越小，峰谷越来越少——那么它的能量可能会越来越低。但是，正如我们在第4章的场合看到的，量子力学告诫我们，这样的推论是错误的。在量子力学看来，弦跟其他所有的振动和波动一样，只能以分离的单位存在。大体上说，一个振动模式所赋予的能量是某个最小能量单元的整数倍，就像那个仓库里的伙伴们拿的钱，都是某个钞票单位的整数倍。特别地，这里说的最小能量单元正比于弦的张力（它也正比于相应振动模式的峰和谷的数目），而那整数则由振动模式的振幅决定。

我们现在讨论的要点是：因为最小能量单元正比于弦的张力，而弦的张力很大，所以，在基本粒子物理学的一般尺度上，这个基本的能量单元也是很大的。它们是所谓普朗克能量的倍数。这个量有多大呢？假如我们用爱因斯坦著名的转换公式$E=mc^2$将普朗克能量化成质量，相应的质量将是质子质量的千亿亿（$10^{19}$）倍。这个以基本粒子的标准看来庞大的质量，就是普朗克质量，大概相当于一粒沙尘或者一百万个细菌的质量。这样，在弦理论图景中，振动的弦圈所对应的典型质量一般是普朗克质量的整数（1，2，3……）倍。关于这一点，物理学家常说，弦理论的“自然”或“典型”的能量尺度（当然也是质量尺度）是普朗克尺度。

这引出一个大问题，直接与我们想再现表1.1和表1.2的粒子性质的愿望有关：如果弦理论“自然”的能量尺度比质子大千亿亿倍，它又如何能解释构成我们周围世界的那些“轻飘飘”的粒子——电子、夸克、光子等等？

问题的答案还是来自量子力学。不确定性原理保证了没有什么东

西是绝对静止的，所有物体都在经历着“量子战栗”，否则我们就会完完全全地知道物体在哪儿，运动多快，那就违背海森堡的原则了。这一点对弦理论中的弦圈也是成立的。一根弦圈，不论显得多宁静，也总是经历着一定的量子振荡。20世纪70年代发现了一件令人惊奇的事情，前面图6.2和图6.3示意的那些直观的弦振动会与量子振荡发生能量的“消减”。就是说，由于量子力学的奇异性，与弦的量子振荡相关联的能量是负的，它将振动弦的总能量减少了大约普朗克尺度的能量。这意味着我们曾天真地以为等于1倍普朗克能量的弦振动模式的最低能量将大大地减少，从而生成相对低能的振动，它们相应的等价质量正好处在表1.1和表1.2的物质粒子和力的信使粒子的质量附近。于是，正是这些最低能量的振动模式，应该能在弦的理论图景和实验能及的粒子物理世界之间建立某种联系。一个重要的例子是，谢尔克和施瓦兹发现，在那个性质像引力的信使粒子的振动模式中，能量彻底地消失了，结果产生一个零质量的引力的粒子，恰好是我们所期待的引力子；因为引力是以光速传播的，而只有零质量的粒子才能以这个极大速度运行。但是，低能振动的组合却更多是例外，而不是一般法则。更典型的振动基本弦所对应的粒子，质量一般要比质子大千百亿亿倍。

这些事实告诉我们，表1.1和表1.2中相比之下轻得多的基本粒子应该是以某种方式从高能量弦的咆哮的朵朵浪花里产生出来的。即使顶夸克那样有189个质子质量的重粒子，也能从振动的弦生成，不过，这需要让弦的巨大的普朗克尺度的特征能量在量子不确定的涨落中减小到原来的一亿亿分之一。这好像在“买得巧”[34]的游戏中，主持人给你一千亿块钱，叫你去把它花了，或者说，把它减少，只留下189块，不能多，也不能少。拿着那么多钱，要花得那么精确，还不知道每样东西的精确价格，即使世界上最精明的买卖人也会大伤脑筋的。在弦理论中，流通的不是钞票，而是能量，近似计算证明了类似的能量消减一定能够出现。不过，要证明这种消减在那么高的精度上实现，总的说来超出了我们今天的理论水平，在随后的几章里我们会

逐渐明白那是为什么。即使这样，我们还是可以看到，像以前说过的那样，弦理论中的许多其他对细节不那么敏感的性质，都能抽象出来，并满怀信心地理解它们。

这将我们引到巨大弦张力的第三个结果。弦能以无限多的不同的振动方式振动，例如在图6.2里我们画了几个峰谷数越来越多的弦振动模式，那才是一个无限序列的开头。这似乎意味着它还对应着一个无限的基本粒子序列，那不是显然与表1.1和表1.2概括的实验情况相矛盾了吗？

是的，的确如此。如果弦理论是对的，无限多弦共振模式的每一个都应该对应一个基本粒子。不过，还有基本的一点，强大的弦张力保证除了几种振动模式（几种能量最低的振动，能量差不多被量子涨落消减净了）而外，其他的都对应着极重的粒子。这里，“重”的意思是，比普朗克质量还重许多倍。我们最强大的粒子加速器所能达到的能量只有质子质量的1000倍，还不及普朗克能量的千亿分之一。所以，在实验室里寻找弦理论预言的那些新粒子，离我们还遥远得很。

然而，我们却有许多间接的办法来寻找那些粒子。例如，在宇宙诞生之初，能量应该是很高的，足以产生大量那样的重粒子。当然，我们一般不会指望它们能留存到今天，因为这些超重的粒子往往是不稳定的，会通过一级一级的衰变失去大质量，最终成为我们熟悉的寻常世界的轻粒子。不过，这些超重的弦振动状态，大爆炸的遗迹，也可能真的会留到现在。毫不夸张地讲，找到这样的粒子可是不朽的发现，在第9章我们会更详细地讨论。

弦理论中的引力和量子力学

弦理论搭建的统一框架是很吸引人的，但它真正吸引人的地方在于它能缓和引力与量子力学间的对立。我们都记得，在结合广义相对论与量子力学时，问题就发生了，那是两个理论的核心特征碰撞的结

果——在广义相对论里空间和时间形成一个光滑弯曲的几何结构；而在量子力学中，宇宙万物，包括空间和时间，都在经历着量子涨落，而且，在越小的距离尺度上，涨落越剧烈。在普朗克尺度以下，疯狂的量子涨落打破了光滑弯曲的几何概念，也就推倒了广义相对论的基础。

弦理论“抹平”了空间的短距离性质，从而令喧嚣的量子波浪安静了许多。这到底是什么意思？它是怎么解决矛盾的？关于这些问题，我们有一个粗略的回答，还有一个更准确的回答，下面就依次来讨论。

粗略的回答

大体上说，我们认识物体结构的一种办法是，用其他事物来打击它，然后观察它们沿着什么路线偏转。例如，我们能看见东西，是因为从那东西反射回来的光子带着信息到达我们的眼睛，然后我们的大脑识别了这些信息。粒子加速器建立在同样的基础上：它让电子和质子等物质相互碰撞，也让它们去撞击其他目标，然后，精密的探测仪器来分析产生的碎末，从而决定那些目标所包含的结构。

一般说来，我们所用的探针粒子的大小决定了我们所能探测的尺度的下限。为认识这句话的重要性，我们来看一个例子。斯里姆和吉姆兄弟想学点儿艺术，于是他们报名进了一个绘画班，经过一段时间的课程，吉姆越来越讨厌斯里姆那一副美术家的派头。他想跟他玩儿一场不同寻常的比赛。他提议每人拿一粒桃核，固定在台钳上，然后画一幅精确的静物图。吉姆的挑战的不同寻常在于谁也不许看着桃核，而是向核发射东西（当然不是光子！），通过观察东西的偏转来确定它的大小、形态和特征，如图6.4。吉姆瞒着斯里姆，在他的枪里填满石弹子（图6.4（a）），而在自己的枪里填满小得多的5毫米塑料弹头（图6.4（b））。两人都开枪发射，比赛开始了。

过一会儿，斯里姆的图画好了，如图6.4（a），通过观察石弹子偏转的轨迹，他发现桃核是表面坚硬的小团东西，不过他也只能知道这么多。石弹子太大了，不可能反映出桃核更细的褶皱结构。当斯里姆看吉姆的画时，惊讶地发现他的画比自己的好（图6.4（b））。不过，看一眼吉姆的枪，他知道自己上当了：吉姆用的小弹头足以反映出由桃核表面的一些大结构所引起的偏转角度。所以，在发射许多5毫米弹头后，吉姆可以看到子弹的偏转的情形，然后画出更细的图。斯里姆不服输，回头用更细小的半毫米弹头填满他的枪，这些小探针粒子足以从核表面的细微褶皱间进出。看它们如何偏转，斯里姆就能画出图6.4（c）的那幅胜利的图画。

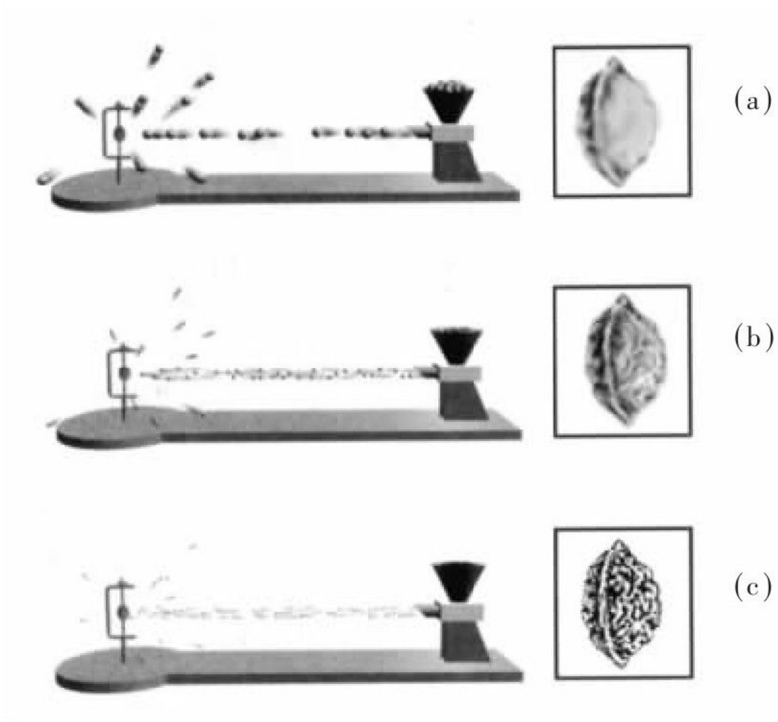



图6.4　桃核固定在架子上，通过观察打在它表面的“探针”的偏转情况来描绘它的图像。所用探头越小——（a）石弹子，（b）5毫米弹头，（c）半毫米弹头——绘出的图像越细致。

这场小小竞赛的教训是很清楚的：我们用的探针粒子不能比所检

验的物理特征的尺度大得太多；否则，它们就感觉不到那些有意义的结构。

假如我们还想更深入地认识桃核的原子和亚原子结构，上面讲的当然还是对的。半毫米的子弹这时不能提供什么信息；它们显然是太大了，不可能对原子尺度的结构产生什么反应。这也是为什么我们在粒子加速器里用质子或电子来作探针的理由，因为它们尺寸小，更适合探测小尺度的结构。在亚原子尺度，量子概念取代了经典逻辑，粒子探针灵敏度的最恰当的度量是它的量子波长，波长表明了它的位置有多大的不确定性。这一点是我们第4章关于海森堡不确定性原理的讨论的结果，在那里我们曾看到，用点粒子做探针（我们主要讲的是光子探针，但讨论也适合于所有其他粒子）引起的误差区间大约等于探针粒子的量子波长。用不那么严格的语言，我们可以说，量子力学的“战栗”把点粒子的探针“抹平”了，就像一位紧张的外科大夫，用颤抖的手拿着手术刀，那开刀的位置还能准确吗？不过，回想一下，我们在第4章还谈到另一点重要事实：粒子的量子波长反比于它的动量，而动量大致也就是它的能量。所以，通过提高点粒子的能量，可以使它的量子波长越来越短——探头越来越“尖”——从而可以用来探测更精细的物理结构。直观地看，高能粒子有更强的穿透能力，所以能探测更细微的特征。

在这一点上，点粒子与弦表现出显著的差别。与塑料弹头探测桃核表面特征的情形一样，弦的空间大小也限制了它不能探测比它自身尺度更小的任何事物的结构——在这里，即那些在普朗克长度以下生成的结构。说得更具体一点，1988年，当时在普林斯顿大学的格罗斯（David Gross）和他的学生孟德（Paul Mende）证明，在考虑量子力学的条件下，持续增大弦的能量并不能持续提高它探测更精细结构的能力，这与点粒子的情形是直接对立的。他们发现，弦能量开始增加时，确实像点粒子那样，能探测更小尺度的结构。但当能量超过普朗克长度下的结构所要求的量时，多余的能量不能使弦探针变得更

尖。相反，那些能量会使弦长大，从而减小它的小尺度灵敏度。实际上，虽然弦的典型尺度是普朗克长度，但是，如果在弦上堆积足够的能量——那是我们怎么也想象不到的大能量（不过，它很可能在大爆炸时出现过）——我们可以使它长大到宏观的尺度，那实际上不可能是灵敏的微观宇宙的探针！看来，弦不同于点粒子，它有两个令探头“迟钝”的根源：一个是量子战栗，与点粒子类似；一个是它自身的空间大小。增大弦的能量可能减小第一个来源的影响，却最终增大了第二个来源的影响。结果，不管我们费多大力气，弦的延伸本性使我们不可能探测普朗克长度以下的现象。

但是，广义相对论与量子力学之间的整个矛盾却出现在普朗克长度以下的空间结构性质。如果宇宙的物质基元不能探测普朗克尺度下的距离，那么不论这些基元还是它们组成的事物，都不可能受那可能的灾难性的小尺度量子涨落的影响。这就像我们用手抚摸一块非常光亮的花岗石，虽然花岗石表面在微观上凹凸不平，是点点细小的颗粒，但我们的手指头摸不出那些细微的变化，只感觉它是完全光滑的。我们粗糙的手指头把小颗粒都“抹平”了。同样，因为弦能在空间生长，它对小尺度的灵敏度也有一定的极限。它“感觉”不出普朗克距离尺度下的变化，它像我们的手指一样，把引力场的超微观涨落都“抹平”了。虽然残留的涨落还很剧烈，但抹平后的光滑已足以平息广义相对论与量子力学的水火不容。而且，还有特别的一点，从引力的量子理论的点粒子方法中产生的那些可恶的无穷大（上一章讨论过了），都被弦理论干净地消除了。

花岗石与我们关心的真正的空间结构之间的根本区别在于，我们有办法让花岗石表面的微观颗粒结构表现出来：不用手指，用更细、更精的探针，就能做到这一点。电子显微镜能识别比百万分之一厘米还小的表面结构，这足以揭示数不清的表面缺陷。相反，在弦理论中，普朗克尺度以下的空间结构“缺陷”是没有办法暴露出来的。在弦理论定律主宰的宇宙中，我们不能再像传统那样把大自然无限地分割

下去。分割是有极限的，在我们遭遇图5.1中吞没一切的量子泡沫之前，那极限就会出现。因此，在某种意义上我们甚至可以说，假想的普朗克尺度下汹涌的量子波浪是不存在的，以后我们还会把这话讲得更准确一些。实证主义者总是认为，只有——至少在原则上——可以探寻和测量的事物才是存在的。因为弦被看做是宇宙最基本的东西，又因为普朗克尺度以下的空间结构涨落的波澜不足以影响这些相对说来巨大的弦，所以，那些涨落是无法测量的，从而在弦理论看来，它们实际上是不存在的。

花招？

上面的讨论可能不会让你满意，我们没有说明弦理论如何克服普朗克尺度以下的空间量子涨落，而是似乎用弦的尺度来回避了整个问题。我们真的解决了什么问题吗？是的。下面讲的两点会让我们更清楚一些。

首先，以上的讨论说明，假想的普朗克尺度以下的空间涨落是在以点粒子框架建立广义相对论和量子力学时产生的人为现象。所以，从某种意义说，当代理论物理学的核心矛盾是我们自己造出来的问题。以前，我们想象所有的物质的粒子和力的粒子都是点状的东西，没有空间大小，所以我们也总觉得要在任意小的空间尺度下考虑宇宙的性质。而在最小的尺度上，我们走进了似乎不可逾越的问题堆。弦理论告诉我们，我们遭遇那些问题只是因为没有真正懂得游戏规则；新规则告诉我们，我们在宇宙中将走近一个距离的终点——那实际上是说，我们传统的距离概念在超微观的宇宙结构中并不是无限适用的。我们想象的可恶的空间涨落现在看来不过是从我们的理论生出来的，而原因是我们不知道那些极限，从而在点粒子路线的引导下走过了物理学实在的边缘。

现在我们看到，广义相对论与量子力学间的矛盾就这样简单地克服了。可能有人会奇怪，为什么过了那么久人们才发觉点粒子不过是

一种理想化的描述，而真实世界的基本粒子确实是有空间大小的。这引出我们要讲的第二点。多年以前，理论物理学的一些伟大的思想家，如泡利、海森堡，狄拉克和费曼，的确提出过大自然的基本组成可能不是一些点，而是一些捉摸不定的“点滴”或者“零碎”。然而，他们和其他一些人发现，很难构造一个理论，其中的物质基元不是点粒子，而且还要满足最基本的物理学原理，如量子力学的几率守恒（因为这一点，宇宙间的事物才不会毫无声息地突然消失），信息传播的光速极限。他们的研究从许多方面一次又一次地证明，如果抛弃点粒子的概念，那两个原理也会被破坏。于是，长期以来，寻找一个以点粒子以外的其他事物为基础的合理的量子理论，似乎是不可能的。弦理论真正动人的地方是，20多年来的艰苦研究表明，尽管弦理论有一些陌生的特征，但它的确满足任何一个合理的物理学理论所要求的性质。而且，还有一点，因为振动的引力子模式，弦理论包括了引力的量子理论。

准确的回答

从前面那个粗略的回答，我们基本明白了为什么弦理论在点粒子理论失败的地方独领风骚。所以，如果你愿意，你可以接着读下一节，而不会失去讨论的逻辑连贯。不过，既然第2章已经讲过了狭义相对论的基本概念，我们现在有可能更精确地说明，弦理论如何平息了疯狂的量子“战栗”。

在这个更准确的回答里，我们还是依据大概回答所依据的中心思想，不过直接在弦的水平上表达。我们将通过较为详细地对比点粒子和弦的探针来回答这个问题。我们会看到，弦的延展特性是如何抹去点粒子探针所得到信息的，从而它又是如何走出当代物理学最核心的超短距离下的困境的。

我们先来考虑，假如点粒子真的存在，它们会如何发生作用，从而如何成为物理学的探针。最基本的相互作用发生在两个运动粒子的

碰撞过程中，这时，两粒子的轨迹会像图6.5那样相交。如果粒子是台球，它们会在碰撞以后发生偏转，走上新的轨道。点粒子的量子场论证明，基本粒子发生碰撞时也会发生类似的事情——粒子散射分离，然后飞向偏转的轨迹——不过细节有些不同罢了。




图6.5　两个粒子的相互作用——它们"轰"地撞在一起，然后沿偏转的轨道离开。

为说得具体简单一些，我们假设一个粒子是电子，另一个是它的反粒子，正电子。物质与反物质发生碰撞时，会湮灭为纯能量，生成光子。[4]为区别新生成的光子的轨道与原来的电子和正电子的轨道，我们遵循传统物理学的约定，把光子的路径画成波浪线。一般说来，光子走过一段距离后会把从原来的电子-正电子对得到的能量放出来，生成另一个电子-正电子对，它们的轨迹如图6.6的右端。两个粒子撞向对方，通过电磁力发生相互作用，最后又出现在偏转的轨道上，这个过程与台球的碰撞过程是相似的。




图6.6　在量子场论里，粒子与它的反粒子会在瞬间湮灭，生成光子。然后，光子生成另一对粒子和反粒子，沿不同的轨道飞离。

我们感兴趣的是相互作用的细节——特别是原来的电子与正电子

发生湮灭产生光子的那一点。以后我们会明白，最核心的事实是，湮灭发生在完全可以确定的一个空间和时间点：标在图6.6的那一点。

当我们走近这些零维的点状物体时，它们实际上是一维的弦，这时会出现什么情况呢？相互作用的基本过程还是一样的，不过碰撞的东西是振动的线圈，如图6.7。如果线圈振动的共振模式适当，它们也可能代表像图6.6那样的电子与正电子的碰撞。只有在走近最微小的距离尺度——比我们今天技术能及的任何事物都小得多的尺度，它们真正的类弦特征才能明显地表现出来。与点粒子情形一样，两根弦发生碰撞，在“闪光”中相互湮灭。那闪光的光子本身也是一根特殊振动的弦。于是，两根弦走过来融合在一起，生成第三根弦，如图6.7。像点粒子的图景那样，新生的弦经过一小段距离，然后释放出原来两根弦的能量，生成两根新的弦，继续走下去。除了最微观的方面，这一切看起来还是像图6.6的点粒子相互作用。

可是，在两种图景间还存在着很重要的差别。我们强调，点粒子相互作用发生在空间和时间的一个可以确定的位置，那是所有观察者都能同意的。而我们应该看到，这在弦相互作用是不对的。关于这一点，我们来看第2章的那两位相对运动的观察者，乔治和格蕾茜会如何描述弦的相互作用。我们将看到，关于两根弦第一次在什么时刻、什么地方相遇，他们会有不同的意见。

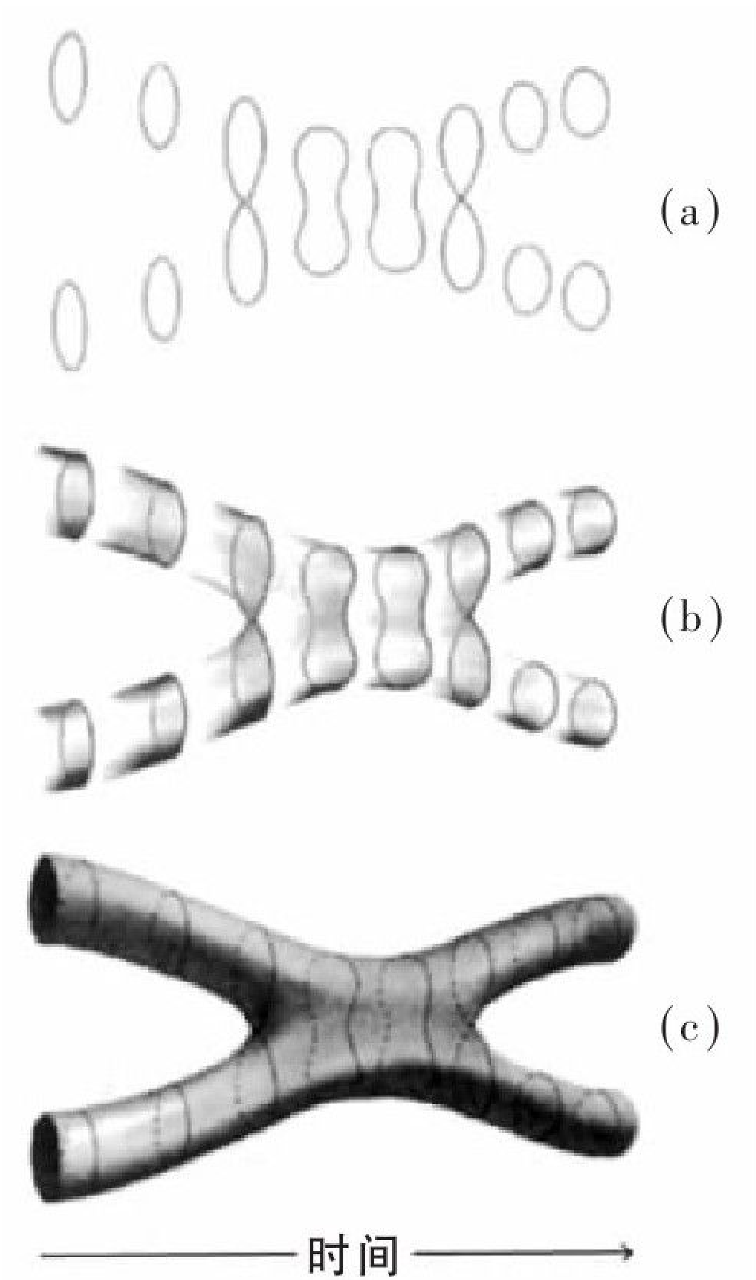



图6.7　（a）两根碰撞的弦可以结合成第三根弦，然后再分裂成两根弦沿偏转的轨道运动下

去。（b）是与（a）相同的过程，强调了弦的运动。（c）两根相互作用的弦随时间流逝而扫过一张“世界叶”。

我们想象用摄像机来观察两根弦的相互作用，把全过程拍成一小段电影，5结果是图6.7（c）的所谓弦的世界叶。把世界叶“切割”成一些相互平行的片——如面包片——我们能恢复弦相互作用的每一瞬间的历史。在图6.8里我们画了切割的例子。具体说，图6.8（a）是乔治看到的事情，他关心的是两根过来的弦；图中还画了一张切割的平面，切过空间所有在他看来同时发生的事件。像往常一样，为了图像更清晰，我们压缩了空间维。实际上，任何观察者看到的同时发生的事件都应该是一个三维的序列。图6.8（b）和图6.8（c）是在稍后时刻的两个镜头——后来的一“片”世界叶——它们说明乔治看到的两根弦是如何靠近的。最重要的是，我们的图6.8（c）定格在两根弦第一次相遇的瞬间（当然是乔治看到的），两弦结合在一起，生成一根新弦。

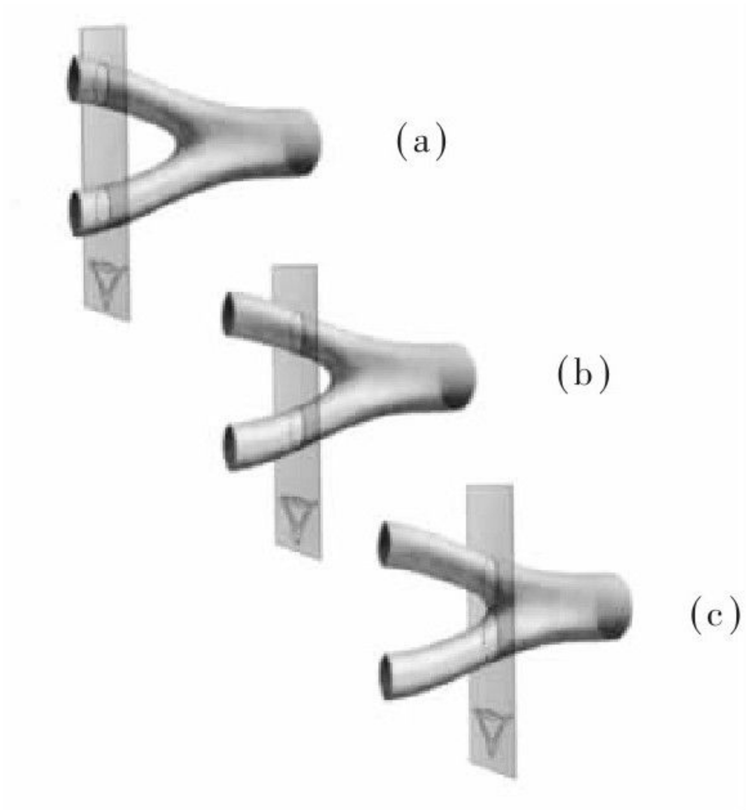



图6.8　乔治看到的两根弦在相继三个时刻的样子。在（a）和（b），两根弦越靠越近；它们在（c）第一次接触（从他的观点看）。

现在来看格蕾茜的情形。我们在第2章讲过，因为格雷茜与乔治是相对运动的，关于事件是不是同时发生，他们会有不同的观点。从格蕾茜的观点看，在空间同时发生的事件处在不同的一张面上，如图6.9。那就是说，在她看来，图6.7（c）的那个世界叶应该以另外的角度切割才能反映相互作用在每一个瞬间的表现。

在图6.9（b）和图6.9（c），我们画了后来两个时刻的情形（现在是从格蕾茜的观点画的），包括她看到两根弦相遇生成第三根弦的瞬间。

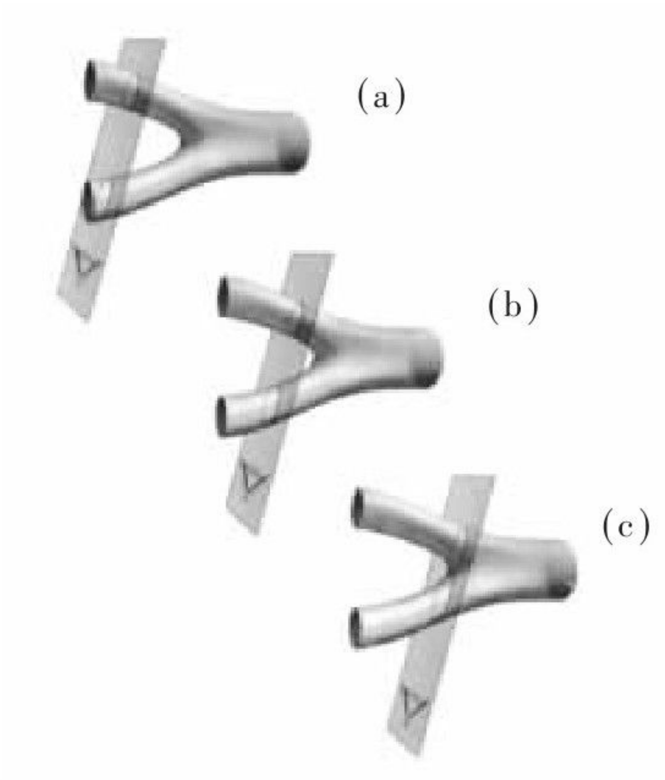



图6.9 格蕾茜看到的两根弦在相继三个时刻的样子。在（a）和（b），两根弦越靠越近；它们在（c）第一次接触（从她的观点看）。

图6.10把图6.8（c）和图6.9（c）放到一起来比较，我们看到，关于原来的两根弦在什么时候、什么地方第一次相遇——发生相互作用，乔治和格蕾茜有不同的意见。因为弦是有空间大小的，它们在空间的什么地方、在什么时刻第一次发生相互作用，不可能有明确的位置——那依赖于观察者的运动状态。

把同样的论证用于点粒子的相互作用，如图6.11，我们还是能得到以前讲过的结论——点粒子的相互作用在确定的时刻发生在空间确定的一点。点粒子把一切相互作用都挤进一个确定的点。当相互作用的力是引力——就是说，传递相互作用的信使粒子是引力子，而不是光子——那么，完全挤在一个点的相互作用将带来灾难性的结果，如

我们以前提到过的无限大结果。反过来，弦把发生相互作用的地方“抹开”了。因为不同观察者看到相互作用发生在图6.10左边不同位置的切面上，相互作用实际上就在所有这些面上展开了。这样，力的包裹打开了，在引力的情形，超微观的“浓缩”性质也大大地淡化了——于是，原来计算无限大的地方，现在出现了很好的有限的结果。这就是我们在前一节大概回答时讲过的“抹平”的准确意思。当然，在普朗克长度距离以下模糊的超微观空间涨落也因此而抹平、光滑了。



图6.10　乔治和格蕾茜看到的发生相互作用的位置是不同的。

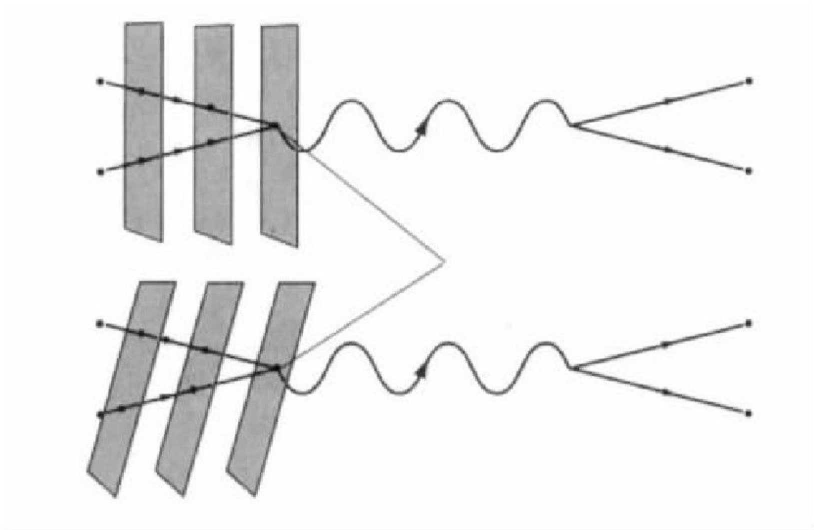


图6.11　相对运动观察者会看到两个点粒子的相互作用在同一时刻发生在空间的同一点。

从弦理论看世界，就像戴着不适当的眼镜看东西，原来点粒子探针能探测到的普朗克尺度下的精细图景，在弦看来成了模糊的一片，不再令人害怕了。不过，弦理论不是近视眼，它看到的就是宇宙的最终图景，不存在校正的透镜去聚焦什么普朗克尺度下的涨落。广义相对论与量子力学的矛盾只有在普朗克尺度下才会明显表现出来，而在距离——传统意义上能够达到或者可能存在的距离——有下限的宇宙中，矛盾是可以避免的。那就是弦理论所描绘的宇宙，在这里，我们看到“大”定律与“小”定律和谐地走到一起了，而过去感觉会在超微观尺度上出现的灾难，则烟消云散了。

弦外

弦有因为两点而奇特。第一点，弦虽然在空间延展，但还是可以很好地在量子力学的框架里描述，第二点，在无数的共振模式中，有一种完全具有引力子的性质，这使得引力成为弦结构的一个天然组成部分。然而，既然弦理论证明了传统的零维点粒子是一种数学的理想化，而不是真实世界的再现，那么无限细小的一维弦圈会不会也是一

种数学理想呢？真实的弦也可能是有粗细的——如二维的自行车胎，或者甚至更“真实”地像三维的面包圈？这条自然路线研究者们从来没有走出结果，那困难似乎是难以逾越的。当年海森堡、狄拉克等人为了构造一个关于三维物质基元的量子力学，也没能走过去。

然而，谁也没想到，在20世纪90年代中期，弦理论家们通过间接但精妙的论证发现，那种高维的物质基元确实在弦理论中扮演着重要而微妙的角色。研究者们逐渐意识到，弦理论并不是只包含了弦的理论。1995年由惠藤等人发动的第二次超弦革命的一个重大发现就是，弦理论实际上还包含着许多不同维的东西；它们像二维的飞盘、三维的小水滴，甚至可能像别的更奇异的怪物。有关的最新认识留到第12章、第13章讲。现在我们还是继续沿着历史的路线，去看看一维弦生成的宇宙比点粒子宇宙，有什么惊人的新性质。

注释

1.标准模型真有一个让粒子获得质量的机制——希格斯机制，是以苏格兰物理学家希格斯（Peter Higgs）的名字命名的。但是就解释粒子质量而言，这不过是把问题转移去解释一种假想的“出让质量”的粒子——所谓希格斯玻色子——的性质。实验正在寻找这种粒子。不过，像我们说的那样，即使粒子找到了，性质测量了，那也是标准模型的输入数据，理论并不能解释它们。

2.为了喜欢数学的读者，我们可以把弦振动模式与力荷的关联描写得更准确一些：弦运动量子化以后，可能的振动状态像在任何量子力学系统中的一样，可以用希尔伯特空间的矢量来表示。这些矢量可以拿它们在一组对易厄米算子下的本征值来标记。算子之一是哈密顿算子，它的本征值是振动态的能量，也就是质量；还有些别的算子，能生成理论需要的不同的规范对称。这些算子的本征值就生成相应的弦振动态所携带的力荷。

3.通过第二次超弦革命（在第12章讨论），惠藤和费米国家加速器实验室的里肯（Joe Lykken，他是更令人瞩目的学者）发现这个结论可能有点儿微妙的问题。考察这些发现后，里肯提出，弦的张力可能会小得多，这样弦就比以前想的大得多。弦大了，我们有可能在下一代粒子加速器里看到它。假如这种可能是真的，那么一个激动人心的前景就会展现在我们眼前——这里和在以后讨论的弦的许多令人惊奇的东西将在未来的10年里得到实验证明。不过，即使弦理论还抱着“更传统的”$10^{-33}$厘米大小的“小”弦，我们还是有很多间接的方法来寻找它们，这将在第9章讨论。

4.专业的读者会发现，在电子-正电子碰撞中产生的光子是虚光子，所以必然会在短时间内“归还”能量，分裂成电子-正电子对。

5.当然，摄像机是在“收集”从物体反弹回来的光子并把光子记录在胶片上。我们在这个例子中用的摄像机不过是一个符号，因为我们并不想看到从碰撞的弦反弹回来的光子。我们只是想在图6.7（c）中记录整个相互作用过程，说明这点以后，我们该指出正文里忽略了的更微妙的一点。第4章讲过，我们可以用费曼的路径求和的办法来建立量子力学，那个方法是，把物体从某个起点到某个终点的所有可能的路线组合起来（每条路线都有一个费曼确定的统计权重）。在图6.6和图6.7里，我们只画了点粒子或弦的从起点走到终点的无数可能路线中的一条。但是这里的讨论同样适用于任何其他可能的路径，从而也就适用于整个量子力学过程。（费曼在路径求和框架下建立的点粒子量子力学，已经由伯克利加利福尼亚大学的曼德尔斯坦（Stanley Mandelstam）和俄罗斯物理学家、现在普林斯顿大学物理系的波里亚科夫（Alexander Polyakov）推广到了弦理论。）

## 第7章　超弦的“超”

1919年，当爱丁顿成功观测了爱因斯坦预言的太阳引起的星光弯曲时，荷兰物理学家洛伦兹（Hendrick　Lorentz）用电报把这好消息

告诉了爱因斯坦。大家看过这封证实广义相对论的电报后，有个学生问爱因斯坦，如果爱丁顿没有在日食中看到预言的星光弯曲，他会怎么想。爱因斯坦回答说，“那我会为亲爱的上帝感到遗憾，因为理论真是正确的。”[35]当然，假如实验没能证明爱因斯坦的预言，广义相对论就不会是正确的，也成不了现代物理学的基石。不过，爱因斯坦的意思是，广义相对论以那么深刻而美妙、简单而有力的概念描写了引力，很难想象大自然会“错过”它。在爱因斯坦看来，广义相对论太美了，几乎不可能是错的。

然而，美学的认识并不是科学进程的裁判。理论的最终判决是看它们如何经历和面对冷酷、严峻的实验事实。不过，这话必须满足非常重要的一个条件。一个理论在形成之初总是不完全的，很难评价实验结果。但物理学家还是必须判断和抉择应该往哪些方向发展他们的部分完成的理论。有些抉择是依靠内在的逻辑一贯性；我们当然要求任何一个合理的理论却避免逻辑的荒谬。另一些抉择依靠我们对定性的实验结果的感觉，看它对不同的理论概念有什么意义；我们感兴趣的理论总该与现实世界的某些事物发生联系。不过，当然还有一种情况，理论物理学家的某些抉择是根据美学趣味做出的——那样的理论具有跟我们经历的世界一样精妙美丽的结构。当然，美的不一定是真的。也许，宇宙的结构本来就不如我们凭经验想象的那样美；也许，我们会发现今天的美学标准应用在陌生的地方还需要重大的修正。但不管怎么说，当我们走进这个陌生的时代，理论描写的那片天地越来越难以靠实验去探索时，物理学家更是特别需要依靠这样的美学来帮助他们避免可能的死胡同。现在看来，美学的方法确实带来了力量和光明。

同艺术一样，对称性也是物理学美的一个重要组成部分。不同的是，物理学中的对称性有非常具体而精确的含义。实际上，根据对称性的精确概念和它们的数学结论，物理学家在过去几十年里建立了一些新奇的理论，在这些理论中，物质粒子和力的信使粒子之间的关联

比我们过去想象的要密切得多。这些理论不仅统一了大自然的力，也统一了物质的基本组成，具有最大可能的对称性，因为这一点，它们被称为超对称的。我们将看到，超弦理论就是在超对称框架下树起的一个例子，它既是第一个，也是登峰造极的一个。

物理学定律的本质

我们想象那样一个宇宙，它的物理学定律像赶时髦似的令人捉摸不定——年年变、月月变、天天变，甚至每时每刻都在变。在这样的世界里，如果生命历程没遭破坏，我们还能生存，但至少可以说，我们永远不可能有瞬间停留的感觉。任何一个简单的行为都像在历险，因为世界在随机变化着，谁也不能靠过去的经验预测未来的结果。

这样的宇宙是物理学家的噩梦。物理学家——当然，还有差不多所有的人——都依靠一个稳定的宇宙：今天的定律在昨天是正确的，在明天仍将是正确的（尽管我们还没能把这些定律都找出来）。当然，假如“定律”能在倏忽间改变，我们还能说它是定律吗？这并不是说宇宙是静止不变的；宇宙当然在变，每一瞬间都在以无限多的方式变。我们说的是，主宰这些变化的定律是固定不变的。你可能会问，我们是否真的知道这一点。实际上，我们不知道。但我们成功描写了从大爆炸后的短暂时刻直到今天的宇宙的无数特征，这使我们相信，即使定律在变化，那变化也非常缓慢。符合我们所有知识的，最简单假定就是，定律是不变的。

现在我们想象另一个宇宙，物理学定律像一些风土人情——一个地方有一个地方的风俗，它们都坚决地拒绝外来影响的融合。在这样的世界里周游，你会像格列弗那样，[36]经历许多意外的奇遇。但从物理学家的观点看，这是另一个魔鬼的世界，在那里生活真是太难了。例如，在一个国家甚至更小的地方成立的定律，到另一个地方就不再成立了。但是，如果定律的本性就是多变，会发生什么事情呢？在那样的世界里，一个地方做的实验可能与其他地方的物理学定律毫不相

干。物理学家们必须在不同的地方重复相同的实验，去发现当地的自然定律。谢天谢地，我们所知道的关于物理学定律的一切，到处都是相同的。世界各处的实验都能用同一组基本的物理学定律来解释。而且，我们还能用一系列不变的物理学原理来解释宇宙中遥远的天体物理学发现的东西，这更令人相信，相同的定律的确处处都是真的。我们从没到过宇宙的另一头，所以我们也不能肯定在别的地方不会有一种全新的物理学在发生作用，但我们还没看到一点儿新物理学的影子。

当然，这并不是说宇宙在不同的地方有相同的样子——或者有相同的具体性质。在月球上踩高跷的宇航员能做许多在地球上做不了的事情，那不过是因为月球的质量比地球小得多，而不是说引力定律从地球到月球有什么改变。牛顿的（或者更准确的爱因斯坦的）引力定律在地球和月球都是一样的。宇航员经历的差别是因为环境条件变了，而不是物理定律变了。

物理学定律不随运用时间和地点而改变，物理学家把这样的性质说成是自然的对称性。物理学家这么讲的意思是，大自然总是平等地——对称地——对待时间的每一瞬间和空间的每个位置，这样就保证了相同的基本定律在大自然发生作用。这些对称性与音乐和艺术中的对称性一样，反映了大自然的秩序与和谐，一样美妙动人。物理学家在说“美”的时候，至少有一部分说的是现象之美——那些从一组简单的普遍定律中产生出来的千姿百态的复杂而多变的现象。

我们在讨论狭义和广义相对论时，还遇到过别的自然对称性。想想相对性原理，那是狭义相对论的核心。它告诉我们，不论观察者以多大的不变速度相对运动，他们的物理学定律都必须是相同的。这也是一种对称性，因为它的意思是大自然平等地——对称地——看待所有的观察者。每一个这样的观察者都有理由认为自己是静止的。当然，这并不是说相对运动的观察者看到的现象都是完全相同的；实际上，正如我们以前讲的，他们各自看到的可能有着许多惊人的差别。

像在地球和月球上踩高跷的人会有不同的经历一样，这些观察的差别也反映条件的不同——观察者在相对运动着——连他们的观察也是相同的定律所决定的。

爱因斯坦通过广义相对论的等效原理把对称性的内容又扩大了许多，物理学定律对所有观察者都是相同的，即使他们在经历着复杂的加速运动。我们还记得，爱因斯坦的等效原理来自他的一个发现：加速的观察者完全有理由说他自己是静止的，而将他所受的力归结为一个引力场。一旦引力走进这个框架，所有可能的观察者的立场就完全平等了。我们已经看到，所有运动一律平等的对称性原理，除了有内在的美学趣味，在爱因斯坦发现的有关引力的奇异结果中，也起着关键的作用。

自然定律可能牵涉到的与时间、空间和运动有关的对称性原理，只要你肯去想，还会遇到更多。例如，物理学定律与观测的角度无关。你可以做一个实验，然后将所有仪器转一个角度再做一次，它们都遵从同样的定律。这就是所谓的旋转对称性，意思是物理学定律认为所有的方向都是平等的。这也是一个与我们前面的讨论一样的对称性原理。

还有什么我们忽略了的对称性吗？你可能会想到我们在第5章讨论过的与非引力作用相关联的规范对称性。那当然也是自然的对称性，不过太抽象了。我们这里只讲那些与时间、空间和运动有直接联系的对称性。这样的话，似乎不会再有别的可能的对称性了。实际上，物理学家科尔曼（Sidney Coleman）和曼都拉（Jeffrey Mandula）在1967年就证明了，除刚才讨论的而外，不会再有别的与空间、时间和运动相关的对称性能生成一个与我们的世界有任何联系的理论。

然而，经过许多物理学家的仔细研究，后来发现这个科尔曼-曼都拉定理有一点微妙的毛病：它没有完全考察与某种叫自旋的东西密切

相关的对称性。

自旋

基本粒子（如电子）能像地球绕太阳旋转那样绕着原子核转动。但在传统的电子的点粒子图景中，似乎没有什么现象对应于地球绕自己轴的自转。物体自转时，转轴上的点——像飞盘的中心点一样——是固定不动的。如果什么东西真的像一个点，那它就不会有什么转轴以外的“其他点”，所以也不会有点粒子自旋的概念。但是，这个论证却因另一个量子力学奇迹而失去了意义。

1925年，荷兰物理学家乌伦贝克（George　Uhlenbeck）和戈德斯米特（Samuel　Goudsmit）发现，与光被原子发射和吸收有关的大量令人困惑的数据都可以通过假定电子具有特别的磁性来解释。大约百年前，法国人安培（André-Marie Ampère）就证明了磁性来自电荷的运动。乌伦贝克和戈德斯米特沿着这条思路发现，只有一种特别的电子运动形式才能产生实验数据所要求的磁性，那是一种特别的转动——即自旋。与传统观念不同，乌伦贝克和戈德斯米特声称，电子有点儿像地球，既公转，也自转。

乌伦贝克和戈德斯米特果真说的是电子在自旋吗？是，也不是。他们的研究所显示的确实是一个量子力学的自旋概念，多少有点儿像寻常的自转，但本质上却是量子力学的。这是一个微观世界的性质，它清除了经典概念，添加了实验证实的量子特征。例如，我们看一位旋转的溜冰者，当她放下手臂时，会转得更快；当她张开手臂时，会转得更慢。但不论她原来转得有多快，她迟早会慢慢停下来。乌伦贝克和戈德斯米特发现的自旋却不是这样的。照他们的实验和后来的研究，宇宙的每一个电子总是永远地以固定不变的速率旋转。电子自旋不是我们习惯的那类物体偶然发生的短暂的旋转运动，而是一种内禀的性质，跟它的质量和电荷一样。如果电子没有自旋，它也就不是电子了。

虽然自旋先是在电子身上发现的，物理学家后来发现这种思想也同样适用于表1.1的那三族物质粒子。这完全是正确的：所有的物质粒子（连同它们的反物质伙伴）都有与电子相同的自旋。用专业的话讲，物理学家说物质粒子有1/2-自旋，这里的1/2大体上代表着粒子旋转快慢的量子力学度量。[37]另外，物理学家还证明，除引力外的那些力的传递者——电磁作用的光子、弱规范玻色子和强作用的胶子——也都有着内禀的自旋特征，是物质粒子的两倍，都是“1-自旋”。

那么，引力呢？对了，在弦理论之前，物理学家就能确定那种假想的引力子应该有多大的自旋才能成为引力的传播者，答案是光子、弱规范玻色子和胶子的两倍——“2-自旋”。

在弦理论背景下，自旋与质量和力荷一样，也关联着弦的振动模式。与点粒子情形一样，这可能会让人错误地以为弦产生的自旋真是因为弦在空间旋转，不过这样的想象的确让我们在头脑里有一个大概的图景。顺便说一下，我们现在可以把以前遇到的一个重要问题说得更清楚一些。1974年，在谢尔克和施瓦兹发现弦理论应该看成一个包含了引力的量子理论时，他们就是那样想的。他们发现，在所有的弦振动模式中，必然有一种是没有质量的2-自旋的——那正是引力子的标志性特征。哪里出现引力子，哪里就有引力。

有了一点自旋概念，现在我们来看上面提到过的问题：自旋是如何暴露科尔曼-曼都拉关于所有可能自然对称性的结论的缺陷的。

超对称与超伙伴

我们强调过，虽然自旋在表面上像旋转的陀螺，但在本质上却是基于量子力学的结果。1925年发现自旋时，也就发现了一种不可能存在于纯经典宇宙的旋转运动。

这就产生下面的问题：寻常的旋转运动可能满足旋转不变的对称

性原理（“物理学将所有的空间方向都看成平等的”），那么，这种更难捉摸的自旋的旋转运动是不是也能产生什么自然规律的可能的对称性呢？到1971年左右，物理学家证明了回答是肯定的。虽然这段故事很复杂，但基本的意思是，对自旋来说，恰好还有一种在数学上可能的自然规律的对称性，那就是所谓的超对称。[1]

超对称没有一个简单直观的图像；我们所能想象的是，时间的移动，位置的转移，方向的改变，速度的变化，但所有这些可能的看得见的改变都跟超对称牵扯不到一起。不过，就像自旋是“带着量子力学色彩的旋转运动”一样，在“空间和时间的量力学扩张”下，从观察的立场说，超对称性还是可以跟变化发生联系。这里引号里的话是很重要的；后面那句的意思不过是说，超对称性大概在什么地方能走进一个更大的对称性原理的框架。[2]不管怎样，虽然超对称的起源不那么好理解，我们还是要来讲一点它最基本的意义——假如自然律体现了这些原理——这要容易把握得多。

20世纪70年代初，物理学家发现，如果宇宙是超对称的，自然粒子必然成对出现，而自旋相差半个单位。这样的粒子对，不论看做点（如标准模型）还是看做振动的小圈，都叫一对超伙伴。因为物质粒子自旋为1/2，而多数信使粒子的自旋为1，这样看来，超对称让物质粒子与力的粒子配成了对，结成了伴。这似乎是一个美妙的统一图景。问题出在一些细节上。

到20世纪70年代中期，当物理学家想让标准模型包容超对称时，他们发现，表1.1和表1.2的那些粒子，没有一个能做另一个的超伙伴。相反，详细的理论分析表明，如果宇宙具有超对称性，那么每一个已知的粒子都必然有一个尚未发现的超伙伴粒子，它的自旋比已知的伙伴小半个单位。例如，电子应该有自旋为0的伙伴，这个假想伙伴的名字叫超电子（超对称电子的简写）。其他物质粒子也该是这样的。例如，中微子和夸克的假想0自旋伙伴叫超中微子和超夸克。类似地，力的粒子应该具有1/2-自旋的超伙伴：光子有光微子

（photino），胶子有胶微子（glui-no），W，　　　Z玻色子有W微子（wino）和Z微子（zino）。

再走近些看，超对称性似乎是一种很“浪费”的特征，它需要一大堆新的粒子，结果把基本粒子的数目加大了一倍。因为这些超伙伴粒子一个也没发现过，你可以把第1章里拉比为μ子说过的那句话说得更干脆些，“没人想要超对称”，而且你可以完全拒绝这个对称性原理。然而，许多物理学家强烈地感到，那么干脆地把超对称性扔了，还为时过早，原因有三点，我们下面就来讨论。

弦理论之前的超对称

第一点，在美学立场上，物理学家觉得很难相信大自然遵从了绝大多数数学可能的对称，却不遵从余下的那些对称。当然，也许实际出现的就是这样不完全的对称，那是很令人遗憾的。仿佛巴赫在用无数相互交织的乐音实现他那天才的对称的乐曲时，忘了最后几个决定性的音节。[38]

第二点，假如理论是超对称的，即使在忽略了引力的标准模型里，与量子过程相关的那些棘手问题也将迎刃而解。基本的问题在于，每一种粒子都是微观的量子“热浪”的一朵浪花。物理学家发现，在这沸腾的量子池塘里，某些粒子相互作用的过程，只有在标准模型里的参数经过精细调节——精确到千万分之一——从而消除了可恶的量子效应以后，才可能没有矛盾。那样高的精度大概相当于用枪去瞄准月亮上的一个目标，而偏差还不能超过一个变形虫的大小。[39]虽然类似的数字精度能在标准模型中实现，但许多物理学家还是怀疑这样的理论——它太敏感了，即使它所依赖的某一个数在小数点后面第15位有一点儿改变，它也会崩溃。[3]

超对称性极大改变了这种状况，因为玻色子——自旋为整数的粒子（以印度物理学家玻色（Satyendra　Bose）的名字命名）——和费

米子——自旋为半整（奇）数的粒子（以意大利物理学家费米（Enrico　Fermi）的名字命名）——可能产生相互抵消的量子力学效应。它们像一块跷跷板的两端，如果玻色子的量子波浪向上，费米子就要将它压下去。因为超对称性保证了玻色子和费米子是成对出现的，所以某些疯狂的量子效应从一开始就基本平息下来了。这样看来，超对称标准模型——添加了所有超对称伙伴粒子的标准模型——的和谐，不再依赖于令人难过的敏感的数字调节。尽管这是一个很技术的问题，许多粒子物理学家还是认为它使超对称性更有吸引力了。

超对称性的第三点间接证据来自大统一的思想。自然界四种力的一个令人疑惑的特征是，它们固有的强度悬殊太大。电磁力不足强力的百分之一，弱作用大概比电磁力还弱一千倍，而引力只是弱力的千亿亿亿亿分之一（$10^{-35}$）。1974年，格拉肖和他在哈佛的同事乔基（Howard　Georgi）根据他本人和萨拉姆、温伯格曾赢得诺贝尔奖的开创性研究，在电磁力、弱力和强力间建立了类似于（我们在第5章讨论过的）电磁力与弱力间的联系。他们提出的引力外的三种力的“大统一”与弱电理论有一点根本的不同：电磁力与弱力是宇宙温度降到一千万亿开尔文（$10^{15}$K）时从更对称的统一中分离出来的，而乔基和格拉肖证明，与强力的统一只有在更高的温度下——约一万亿亿亿开尔文（$10^{28}$K）才是显著的。从能量看，这相当于质子质量的一千万亿倍，或者说，大约比普朗克质量小四个数量级。乔基和格拉肖大胆地把理论物理学领进了一个高能量的领域，比过去人们所能探索的能量高出好多个数量级。

同一年里，乔基、奎恩（Helen　　Quinn）和温伯格在哈佛的研究，将三种力的潜在统一性在大统一的框架下更显著地揭示出来了。他们的成果对力的统一和超对称性与自然界的关系的评判起着重要作用，所以我们花点儿工夫来解释它。

我们都知道，两个带相反电荷的粒子的电吸引力和两个有重物体间的万有引力随着物体间距离的减小而增强，这是经典物理学里众所

周知的简单特性。但是，当我们研究量子物理学对力的强度的影响时，就会出现一点奇怪的东西。那么，为什么会有量子力学的影响呢？答案还是在量子涨落。例如，当我们考察一个电子的电力场时，我们实际上是隔着一团“云雾”看它——那是在电子周围空间随处出现的瞬间的电子-正电子生成和湮灭形成的“雾”。物理学家先前就发现，这团热腾腾的云雾一般的微观涨落会使电子的力变得模糊，仿佛隔着薄雾看远处的灯塔。不过请注意，当我们走近电子时，一定穿过了那层遮在眼前的粒子-反粒子云雾，从而不太能感觉它们逐渐消失的影响，这意味着，电子的电场强度随我们的靠近而增强了。

物理学家认为，当我们靠近电子时，电场强度的量子力学的增加，根本不同于我们熟悉的它在经典物理学中的增加；量子力学的增加，是因为电磁力的内禀强度随距离减小而增加。这说明，力的增强不仅是因为我们离电子近了，而且还因为我们看到了更多的电子的内禀电场。其实，虽然我们一直在说电子，这些讨论也同样适用于其他带电粒子。总之，我们可以说，在越小的距离尺度上，量子效应使电磁力变得越强。

标准模型里的其他力呢？它们的内禀强度如何随距离改变？1973年，普林斯顿的格罗斯和威切克（Frank Wilczek），哈佛的波利泽尔（David Politzer）分别独立研究了这个问题，发现一个令人惊奇的答案：粒子生成与湮灭的量子云把强力和弱力的强度放大了。就是说，如果我们穿过这团沸腾的量子云，在更近的距离来看这些力时，它们还没经历那样的放大作用。因此，从近距离看，强力和弱力减弱了。

乔基、奎恩和温伯格凭着这点认识，发现了一个重要的事实。他们证明，当这些沸腾的量子效应都仔细考虑进来时，结果是引力而外的三种力将走到一起来。他们认为，这些在当前技术所及的尺度上迥然不同的力，实际上是微观的量子薄雾所产生的不同影响的结果。他们的计算表明，如果不是在寻常尺度上，而是穿过云雾，在十万亿亿亿分之一厘米（$10^{-29}$厘米，只是普朗克长度的一万倍）的距离看这三

种力，它们的强度会变得完全相同。

当然，那个尺度离我们寻常的经验是很遥远的，不过，感应这么小尺度所必需的能量却是混沌、热烈的早期宇宙所特有的——那是在大爆炸后千万亿亿亿亿分之一（$10^{-39}$）秒的时候，我们曾说过，那时宇宙的温度是$10^{28}$K。就像千差万别的物质——如铁、木头、岩石、矿物等——在足够的高温下熔化，形成均匀的等离子体一样，理论研究表明，强力、弱力和电磁力在那样的高温下也会融合成一个“大统一”力。这一点简单地画在图7.1。[4]

虽然我们的技术还不能深入这样小的距离尺度，也产生不了那么炽热的温度，但实验家们自1974年以来已经在日常条件下把那三种力的测量强度大大精确化了。这些数据（图7.1的三条力度曲线的出发点）是乔基、奎恩和温伯格的量子力学外推的前提。1991年，欧洲核子中心（CERN）的阿马尔蒂（Ugo Amaldi）、德国卡尔斯鲁厄（Karlsruhe）大学的德波耳（Wim de Boer）和弗尔斯特瑙（Hermann Fürstenau）用这些数据重做了乔基三人的计算，发现了两样重要的东西。第一，引力外的三种力在微小距离尺度（也就是高能/高温状态）几乎是一致的，但并不完全相同，如图7.2。第二，假如有超对称性，这小小的然而确定不疑的力的偏差就会自动消失。原因是，超对称性需要的新的超伙伴粒子会产生新的量子涨落，这些涨落正好能使那些力的强度趋于一点。

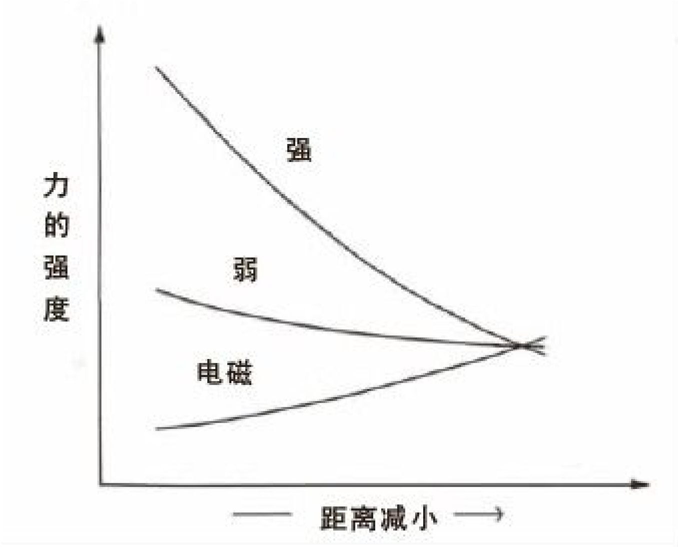



图7.1　引力外的三种力随距离尺度减小——或者说，随能量增加——的作用情况。

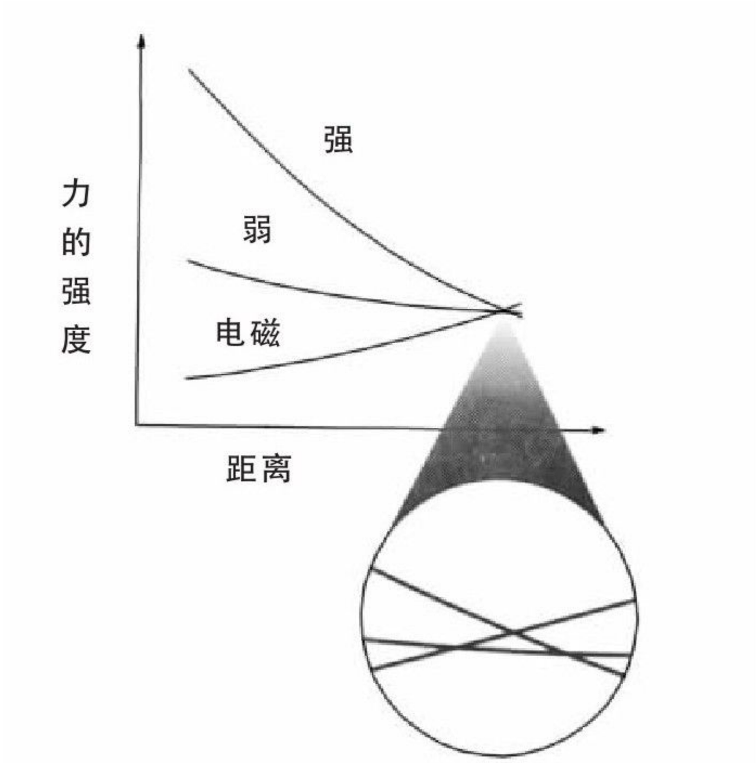



图7.2　力的强度的更精确计算表明，如果没有超对称性，三种力不会完全趋于一点。

大多数物理学家都感到这太难以置信了：大自然竟会这样来选择力——让它们在微观尺度上几乎具有统一的强度（在微观上相等），却还留下一点儿偏差。这就像玩儿拼图游戏时，最后留下一块图板，总不能很好地放进它应该去的地方。超对称性灵巧地把那块图板的形状修正了一点儿，于是可以恰到好处地还原。

最后这个发现的另一点意义是，它为下面的问题提供了一个可能的答案：为什么我们没有发现任何超伙伴粒子？刚才讲的将三种力融合的计算以及许多物理学家研究过的其他问题都表明，超伙伴粒子一定比已知的粒子重很多。尽管还不能有确定的预言，但我们大概知道，超伙伴粒子的质量可能是质子的1000倍（假如不是更重的话）。我们人工的加速器不可能达到这样的能量，所以这也就解释了我们为

什么还没有发现一个这样的粒子。在第9章，我们会回来讨论实验的前景，也许在不远的将来，它们可以决定超对称性是否真的是我们宇宙的一种性质。

当然，让人们相信——至少不拒绝——超对称性，理由还不是那么充分有力。我们讲过，超对称性如何能将理论提高到最大的对称形式，但你可能会说，宇宙本不在乎这些数学独有的最大对称形式；我们讲过，超对称性如何让我们摆脱标准模型里为避免量子问题而调节参数的困难，但你可能会说，真的自然理论也可能就在自我破坏与自我协调间走钢丝；我们讲过，超对称性如何修正了引力外的三种力在小距离的内禀强度，使它们能融合成一个大统一的力，但你还是可能会说，在大自然的设计中，似乎没有什么东西说明这些力应该在微观尺度上相同。而且，最后你可能会说，我们为什么还没找到一个超伙伴粒子，最简单的答案是，宇宙不是超对称的，超伙伴并不存在。

没人能反驳这些回答。不过，当我们考虑超对称在弦理论中的作用时，它就显得力大无比了。

## 弦理论中的超对称

20世纪60年代从维尼齐亚诺的研究中生出的弦理论包括了本章开头讲的所有对称性，但不包括超对称性（那时还没发现呢）。以弦概念为基础的第一个理论，更准确地该叫玻色子弦理论。玻色子的意思是，弦的所有振动模式都具有整数自旋——没有半整数的自旋模式，也就是弦没有费米子的振动模式。这带来两个问题。

首先，如果要拿弦理论来描述所有的力和物质，就必须想办法让它把费米子振动模式也包括进来，因为我们知道物质的粒子都是1/2-自旋的。第二点，也是更令人困惑的一点，在玻色子弦理论中，有一种振动模式的质量（更准确说是质量的平方）是负的——即所谓的快子。虽然在弦理论以前，物理学家就研究过，在我们熟悉的正质量粒

子外还可能存在快子，但他们也发现那样的理论在逻辑上很难（几乎不可能）是合理的。同样，在玻色子弦理论背景下，物理学家为了使奇异的快子振动模式的预言变得合理，曾探讨过各种可能的框架，结果都失败了。这些特点使人们越来越明白，玻色子弦理论虽然很有趣，但一定还存在某些根本性的错误。

1971年，佛罗里达大学的拉蒙（Pierre Ramond）担起了修正玻色子弦理论以囊括费米子振动模式的挑战。经过他和后来施瓦兹和内弗（André Neveu）的研究结果，弦理论出现了新面目。令人惊讶的是，在新理论中，玻色子和费米子的振动模式是成对产生的。每一个玻色子对应着一个费米子，每一个费米子也对应着一个玻色子。到1977年，特林大学的格里奥茨（Ferdinando Glioz-zi）、帝国学院的谢尔克和奥利弗（Dayvid Olive）才发现这些成对出现的粒子的正确意义。新的弦理论包含了超对称性，而看到的这些成对出现的玻色子和费米子振动模式就反映了这种高度对称的性质。超对称弦理论——即超弦理论——就这样诞生了。而且，他们三人还有另一个重要结果：他们证明玻色子弦那令人困惑的快子振动不会损害超对称的弦。这样，一点点的弦困惑慢慢地消失了。

不过，拉蒙、内弗和施瓦兹的研究的最初影响并不在弦理论。到1973年的时候，物理学家韦斯（Julius Wess）和朱米诺（Bruno Zumino）发现，超对称性——从新构造的弦理论中出现的那种新的对称性——甚至也能用于以点粒子为基础的理论。他们很快就迈出重要一步，把超对称引进点粒子的量子场论框架。那时候，量子场论是主流粒子物理学家们的核心——而弦理论正慢慢成为它边缘的一个课题——所以，韦斯和朱米诺的发现所激发的大量后来的研究都集中在所谓的超对称量子场论。上一节讲过的超对称标准模型就是这些探索的一个辉煌成果。我们现在看到，在崎岖的历史征途上，点粒子理论也从弦理论获得过巨大的帮助。

随着超弦理论在20世纪80年代中叶的复兴，超对称性又在原来发

现它的背景下出现了。在这个框架下，超对称性的表现远远超过了上一节讲的。弦理论是我们知道的惟一能融合广义相对论和量子力学的方式，但只有超对称的弦理论才能避免快子问题，才能包括费米子振动模式从而才能说明组成我们世界的物质粒子。为了实现引力的量子理论，也为了一切力和物质的大统一，超对称性与弦理论手拉手地走来了。假如弦理论是对的，物理学家希望超对称性也是对的。

然而，到20世纪90年代中叶，超对称弦理论遇上了一个特别麻烦的问题。

“多”的烦恼

如果有人告诉你，他们解决了埃尔哈特（Amelia Earhart）的失踪之谜[40]，你开始可能感到怀疑；但如果他们有确凿的证据和想好的一套解释，你大概会听他们说下去，说不定还会相信他们。可是接下来，他们告诉你还有一种解释。你也耐着性子听了，惊奇地发现这种解释跟头一个解释一样有根据。这时候，他们又向你讲了第三种、第四种甚至第五种解释——每一种都不同，但都同样令人信服。最后，你一定觉得对埃尔哈特之谜还是跟从前一样，什么也不知道。对一个事物的基本事实解释越多，所知越少，多也就等于无。

到1985年的时候，弦理论——尽管理所当然地激发了许多人的热情——开始有点儿像我们那些过分热心的埃尔哈特专家了。原来，物理学家发现，超对称性（那时已成为弦理论结构的核心元素）实际上可以通过5种不同的方式进入弦理论。每一种方式都能生成成对的玻色子和费米子振动模式，但这些粒子对的具体性质和理论的许多其他性质都有着巨大的不同。尽管名字并不重要，我们还是应该记住这些理论：I型理论，IIA型理论，IIB型理论，杂化O（32）型理论和杂化$E_8 \times E_8$理论。我们讨论过的弦理论的一切特征在这些理论也都能表现出来——只是细节有所不同。

一个包罗万象的理论——一个可能的最终的统一理论——有五种不同的形式，这令弦理论家烦恼。不论埃尔哈特出了什么事情，真正的解释只能有一个（不论我们是否能发现它）；同样，我们希望关于宇宙的最深刻、最基本的认识也应该是这样的。我们生活在一个宇宙，我们希望一个解释。

关于这个问题，一个可能的解决办法是，虽然有5个不同的超弦理论，但其中的四个可以简单地通过实验来排除，最后留下一个真正的相关的解释框架。不过，即使真是那样，我们还是有一个头疼的问题：为什么开始会有那几个理论呢？用惠藤的话来说，“如果5个理论有一个描写了我们的宇宙，那么谁住在其他4个宇宙呢？”[41]物理学家总是梦想寻求最终的答案，引向一个惟一的绝对不可避免的结论。理想地说，最终的理论——不论是弦理论还是其他什么理论——都应该是这样的，不会有别的可能，而只能是它自己。假如我们能发现只有一个逻辑合理的理论能融合相对论和量子力学的基本结构，许多人会认为我们将获得一个对宇宙性质的彻底认识。一句话，那就是大统一理论的天堂。[42]

我们将在12章看到，最近的研究将超弦理论推进了一大步，离统一的乌托邦更近了；那5个不同的理论，原来是描绘同一个宏大理论的5种不同的方法。超弦理论确实有惟一的根源。

问题似乎解决了，但从下一章的讨论我们会看到，通过弦理论走向统一还要求我们离开传统智慧走得更远。

注释

1.超对称性的发现和发展有着复杂的历史。除了文中提到的以外，早期的主要贡献者还有R.Hang, M.Sohnius, J.T.Lapuszanski, Y.A.Gol’fand, E.P.Lichtman, J.L.Gerrais, B.Sakita, V.P.Akulov, D.Y.Volkov, V.A.Sorota，等等。他们的一些工作编辑在Rosanne Di

Stefano, Notes on the Conceptual Development of Supersymmetry, Institute for Theortical Physics, State University of New York at Stony Brook, preprint ITP-SB-8878.

2.对数学感兴趣的读者会看到，这里的推广是在我们熟悉的时空的笛卡儿坐标上添加新的量子坐标，例如u和v，满足反对易关系：u×v＝-v×u。这样，超对称性可以认为是在经过量子力学扩张的时空形式下的一种变换。

3.我们为对具体细节和技术要点感兴趣的读者再多讲几句。在第6章注释1中，我们提到标准模型借助一种“出让质量的粒子”——希格斯玻色子——来为表1.1和表1.2的粒子赋予观察到的质量。为实现这个过程，希格斯粒子本身不能太重；研究表明它的质量不能比质子质量的1000倍更大。但后来发现量子涨落可能为希格斯粒子带来巨大的质量，把它推向普朗克质量的尺度。不过，理论家们发现，这个暴露了标准模型严重缺陷的结果是可以避免的，只要我们把标准模型里的某些参数（特别是所谓的希格斯粒子的裸质量）适当做$10^{15}$分之一的调整，就能消除量子涨落对希格斯粒子质量的影响。

4.图7.1有一点细微的地方需要注意：图中所示的弱力介于强力和电磁力之间，而我们讲过它比那两种力都弱。原因在于表1.2，我们看到，弱力的信使粒子质量很大，而强力和电磁力的信使粒子是没有质量的。本质上说，弱力的强度（用耦合常数来度量，我们在第12章再讨论）是图7.1的样子，不过由于传递粒子活动太慢，所以减小了实际的作用。在第14章我们还将看到引力如何走进图7.1。

## 第8章　看不见的维

爱因斯坦通过狭义相对论和广义相对论，解决了他过去百年的两大科学冲突。尽管从激发他研究的原始问题看不出后来的结果，但两个问题的解决完全改变了我们对空间和时间的认识。弦理论解决了一

百年来的另一个科学冲突，解决的方式很可能连爱因斯坦都觉得惊奇，它要我们的空间和时间的概念经历一个更剧烈的变革。弦理论彻底动摇现代物理学的基础，甚至宇宙的维数——那个我们认为不是问题的基数，也正发生着戏剧性的而且令人信服的改变。

习惯的错觉

经验产生直觉。但经验的作用不止于此：它还为我们分析和解释我们感觉的事物树立一个框架。例如，你一定相信，一群狼养大的“野孩子”会根据与你全然不同的观点来解释世界。即使不那么极端的例子，拿在不同文化传统里成长起来的人来比较，我们也能看到，经验在很大程度上决定了我们认识世界的思想倾向。

当然，有些事情是我们都共同经历过的。往往就是来自这些共同经历的信念和希望，我们最难说得明白，也最难向它们挑战。我们来看一个简单却深刻的例子。假如你放下这本书，站起来，你可以在3个独立的方向——也就是3个独立的空间维——运动。当然，你走任何一条路径，不论多么复杂，都是在3个不同方向的运动的组合——我们一般称那些方向为“左右”、“前后”和“上下”。你每迈出一步，都在做一种选择，决定你如何穿过那3个维度。

还有一种等价的说法，我们在讨论狭义相对论时见过，那就是，宇宙间的任何一个位置都可以用3个数来完全确定：3个数相应于3个空间维。例如，用寻常的话说，城里的某个地址可以用街道（“左右”位置）、路口（“前后”位置）和楼层（“上下”位置）来确定。从更现代的观点说，我们已经看到，爱因斯坦的理论鼓励我们把时间看做另一个维（“过去-未来”维），这样，我们一共有了四维（3个空间维和1个时间维）。为确定宇宙的一个事件，我们应该说它发生在什么时候、什么地方。

宇宙的这个特征是基本的、一贯的，也是普遍存在的，而且似乎

根本成不了什么问题。然而，在1919年，一个无名的波兰数学家，来自柯尼斯堡大学的卡鲁扎（Theodor Kaluza）却敢向显然的事实挑战——他提出，宇宙也许不只有3个空间维，而是有更多。有时候，听起来傻乎乎的话本就是傻话，但也有时候，傻话却动摇了物理学的基础。当然，很久以后我们才会认识到，卡鲁扎的建议变革了我们物理学定律的体系。我们至今还为他的远见感到震惊。

卡鲁扎的理论和克莱茵的改进

宇宙空间不是三维的，可能还有更多维，这话听起来很荒唐，很奇怪，还有点儿神秘。不过，实际看来，那是很具体实在的，也是完全合理的。为看清这一点，我们暂时把目光从浩瀚的宇宙转向我们更熟悉的花园，看一根细长的浇水管。

想象一根几百英尺长的水管横过一道峡谷，从几百米外看，就像图8.1（a）的样子。在这么远的距离上，你很容易看到水管是一根长长的展开的线，如果没有特别好的视力，你很难判断它有多粗。从远处看，如果一只蚂蚁在水管上，你想它只能在一个方向，即顺着水管方向爬行。谁问你某一时刻蚂蚁的位置，你只需要告诉他一个数：蚂蚁离水管左端（或右端）的距离。这个例子的要点是，从几百米以外看，长长的一根水管就像是一维的东西。

实际上我们知道水管是有粗细的。从几百米以外你可能不容易看清，但拿一只双筒望远镜，你可以看得很真切，原来水管是图8.1（b）的样子。在望远镜的镜头里，你还看到有只蚂蚁爬在管子上，能朝两个方向爬行。它可以顺着管子，左右爬行，这一点我们已经知道了；它还可以绕着管子，沿顺时针或反时针方向爬行。现在你明白，为确定某一时刻小蚂蚁在哪儿，你必须告诉两个数：它在管子的什么长度以及它在管圈的什么地方。这说明水管的表面是二维的。[1]

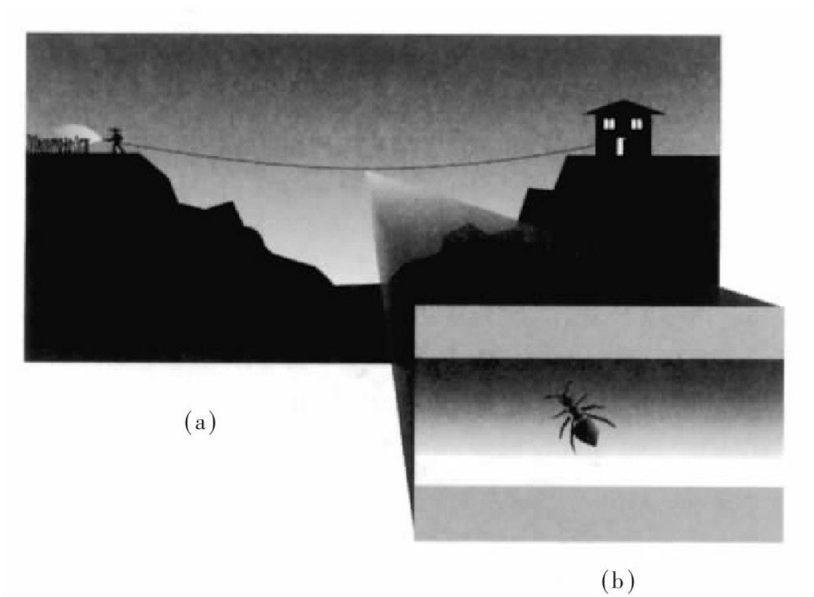



图8.1　（a）从远处看，花园的浇水管就像是一维的。（b）走近来看，水管的第二维就显现出来了——管壁上环绕管道的那一维。

不过，那两维却有很明显的不同。沿着管子伸展方向的一维很长，容易看到，绕着管子的那一圈很短，“卷缩起来了”，不容易发现。为看清圆圈的那一维，你得用更高的精度来看这根管子。

这个例子强调了空间维的一点微妙而重要的特征：空间维有两种。它可能很大，延伸远，能直接显露出来；它也可能很小，卷缩了，很难看出来。当然，在这个例子里你用不着费多大力气就能把“卷缩起来的”绕管子的小圆圈儿揭露出来，那只需要一副望远镜就行了。不过，假如管子很细——像一根头发丝儿或毛细管——要看清那卷缩的维就不那么容易了。

卡鲁扎在1919年给爱因斯坦的信中，提出一个惊人的建议。他指出，宇宙的空间结构可能不只有我们寻常感觉的三维。我们马上就会讨论他提出这一激进问题的动力。原来，他发现这可以提供一个美妙动人的框架，把爱因斯坦的广义相对论和麦克斯韦的电磁理论编织进

单独一个统一的概念体系。但更直接的问题却是，这个建议如何能与我们只看到三个空间维这一显然的事实相协调呢？

问题的答案隐含，在卡鲁扎的理论中；后来，在1926年，瑞典数学家克莱茵（Oskar Klein）把它说得更具体和明确，那就是：我们宇宙的空间结构既有延展的维，也有卷缩的维。就是说，我们的宇宙有像水管在水平方向延伸的、大的、容易看到的维——我们寻常经历的三维；也有像水管在横向上的圆圈那样的卷缩的维——这些多余的维紧紧卷缩在一个微小的空间，即使用我们最精密的实验仪器也远不能探测它们。

为了更清楚地认识这个不同寻常的图像，我们再来看看花园里的浇水管。我们这回绕着管子密密地画满圆圈。同以前一样，从远处看，管子是一根长长的一维的细线。但是，如果拿望远镜来看，很容易看到卷缩的那一维，画了圆圈就看得更清楚了，如图8.2。这幅图说明水管的表面是二维的，1个大的延伸的维和1个小的卷缩的维。卡鲁扎和克莱茵认为，我们的宇宙空间也像这样，不过它有3个大的延伸的维，1个小的卷缩的维——一共是四维。那么多维的东西不好画，为了看得清楚，我们只好将就看两个大维和一个小维的图。图8.3是一个示意图，我们在图中把空间结构放大了，就像用望远镜看水管那样。

图中最下面的一级表现了我们熟悉的周围世界的寻常距离尺度（如若干米）的空间结构，这些距离用大网格表示。接下来，我们关注越来越小的区域，把它放大来看。先看小一点儿的距离尺度下的空间结构，没有什么异常发生；它似乎与原来尺度的结构一样——经过三级放大，我们看到的情景都是这样。不过，当我们在最微观的水平——图8.3的第四级——看空间时，一个新的卷缩的维度出现了，像精心织成的地毯上一个个毛绒绒的小线圈儿。卡鲁扎和克莱茵认为，这些小圈存在于延伸维的每一点，就像水平延伸的水管上处处绕着横向的圆圈。（为看得清楚，我们只在延展的方向上按一定间隔画了些

圆圈的维。）在图8.4里，我们画了一个特写镜头来表现卡鲁扎和克莱茵眼中的空间的微观结构。

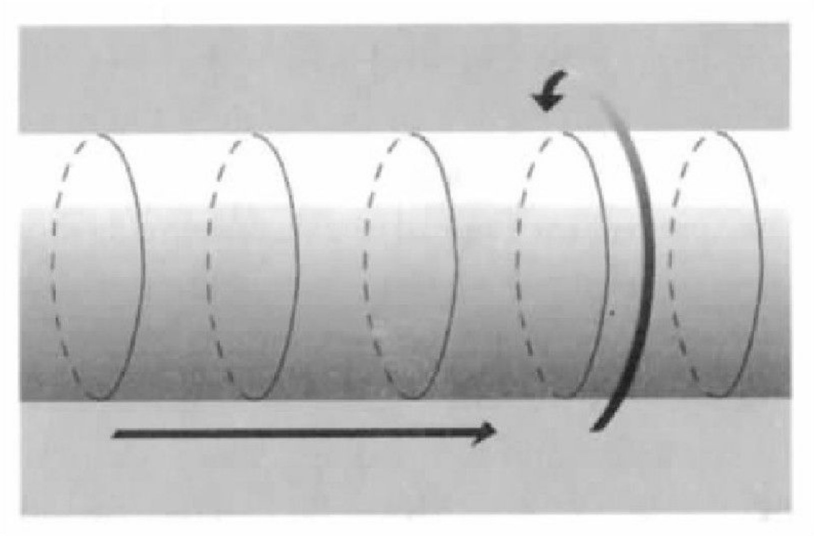


图8.2　花园里浇水的管子是二维的：水平方向的一维由直线箭头表示，是延伸的；横向的一维（圆圈表示）是卷缩的。

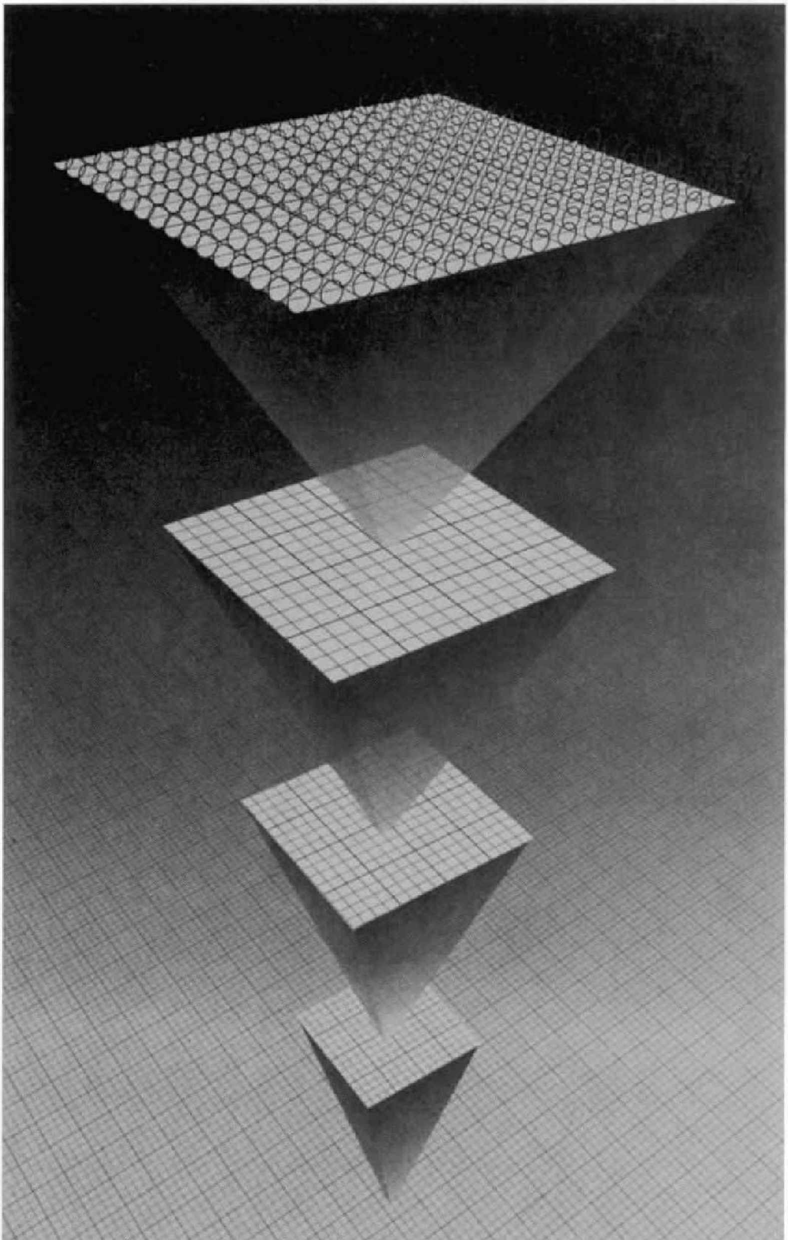


图8.3　类似于图5.1。上一层是下一层表现的空间结构的放大。我们的宇宙可能有额外的维度

——如在第四层看到的——不过它们卷缩在很小的空间里，还没有直接表现出来。

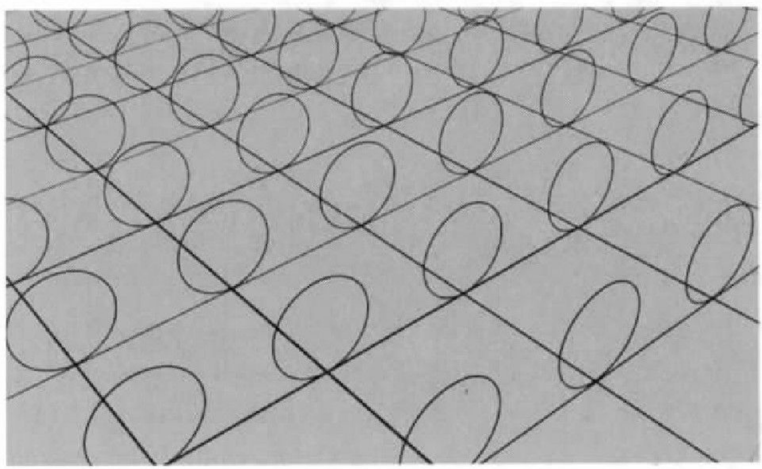


图8.4　网线代表寻常经历的延展维，圆圈代表新的微小的卷缩维。这些圆圈像地毯上的绒毛线圈儿一样，存在于延展方向的每一点——为清楚起见，我们只是把它们画在网格的交点处。

宇宙空间与花园的浇水管子虽然大不相同，但也表现出相似的地方。宇宙有3个大的延展的空间维（我们实际只画了两个），而水管只有一个；更重要的是，我们现在描绘的是宇宙自身的空间结构，不是水管那样存在其间的东西。但是，基本思想是一样的：假如宇宙另一个卷缩的维也像水管的细圆圈儿那样很小，它就会比那些显然的延伸的维难测得多。实际上，如果它太小了，我们用最大的放大器也看不到。另外，最重要的是，这些卷缩的维并不像图上画的那样（你也可能会那么想）是长在延伸方向上的一圈圈“肉瘤”，而是一个新的维度，存在于我们熟悉的空间维的每一点，正如空间每一点都有上下、左右、前后方向一样。这是一个新的独立的方向，蚂蚁（如果足够小的话）可以朝这个方向爬行。为了确定那样一只微观蚂蚁的空间位置，我们不仅需要告诉它在延伸的什么方向（由网格表示），还要告诉在圆圈的什么地方。一个空间位置需要4个数；如果加上时间，我们就得到一条5个数表达的时空信息——比我们平常想的多1个。

这样，我们看到一个令人惊讶的事实：虽然我们知道宇宙只有3个延展的空间维，但卡鲁扎和克莱茵的论证却说明，那并不排除还存在别的卷缩维（至少，如果那些维很小，就是可能的）。宇宙很可能有我们看不见的维。

那些看不见的维多小才算“小”呢？我们最先进的仪器能探测小到百亿亿分之一米的结构。如果那些维卷缩得比这个尺度还小，我们就看不见了。1926年，克莱茵结合了卡鲁扎的原始想法和新出现的量子力学思想。他计算的结果表明，卷缩的维可能小到普朗克长度，是实验远远不可能达到的。从此以后，物理学家把这种可能存在额外小空间维的思想称为卡鲁扎-克莱茵理论。[2]

水管世界的生命

现实的花园浇水管的例子和图8.3的示意，让我们多少能感觉宇宙也可能有更多的空间维。但是，即使这个领域里的研究者，也很难具体“看见”三维以上的宇宙空间。因为这一点，物理学家常常像阿伯特（Edwin Abbott）在1884年的那本迷人的经典流行作品《平直的世界》里描写的那样，[43]想象我们生活在一个维数较低的宇宙，然后逐渐认识宇宙还有我们不能直接感知的更多的维——通过这些想象，我们也养成了对多余维的直觉。现在，我们想象一个二维的宇宙，形状像那花园的浇水管。为此，我们必须抛开“旁观者”的念头，我们不像以前那样“从外面”看宇宙里的一根水管；我们必须忘记原来的世界是什么样的，而走进一个新的管状的宇宙——一根长长的（可以认为无限长）水管的表面就是这个宇宙空间的全部。现在，我们是生活在这个面上的小蚂蚁。

先来看一个有点儿极端的情形。设想管子宇宙很细，细得没有哪个管子上的居民能感觉它的存在。这样，我们生在这个管子宇宙的人们当然相信这样一个基本事实：宇宙空间是一维的。（如果管子世界生出一个小爱因斯坦，他会告诉我们宇宙有一个空间维和一个时间

维。）这个事实如此明显，看来不会有什么问题，于是，我们说自己的家园是“直线国”，就是为了强调它只有一个空间维。

直线国里的生命跟我们所了解的生命大不一样。例如，我们熟悉的身体就不可能适合生活在直线国里。不论你的身体怎么改变，它总是有长度、宽度和厚度——三维的空间延展，这是不可能克服的。直线国没有为这样精美的生命形态留下生存的空间。请记住，虽然在你头脑中直线国可能仍然是存在于我们宇宙空间的一根长长的丝线一样的东西，但是你得把它作为一个宇宙——它就是全部。生活在这样一个家园，你就得适应它那一个空间维。好好想想，即使你像一只蚂蚁，也不能走进它；你必须先变成一条虫子，然后拉得长长的，直到完全失去粗细的感觉。为了生活在直线国，你必须那样，只有长度。

你身体两端各有一只眼睛——那可不像你做人时的眼睛，能在三维空间里向四面张望；直线形生命的眼睛是永远固定的，每一只都盯着前面一维的距离。这并不是你的眼睛长得有问题，你和国中所有的人都知道，那是因为直线国只有一个维，你们的眼睛没有别的方向可以看。直线国的方向只能向前或者向后。

我们还可以进一步想象一些直线国里的事情，但很快会发现那没有多大意义。例如，在你身旁有另一个线形生命，将出现下面的情景：你能看到她的一只眼睛——朝着你的那一只——但不像人眼，而只是一个点。直线上的眼睛没有形状，也没有表情——因为没有它表现那些我们熟悉的特征的余地。而且，你将永远盯着邻居那点一般的眼睛。如果你想探索她身体另一边的直线世界，你会大为失望的。你不可能经过她，她把路“塞满了”，直线国里没有能绕过她的路。当生命在直线国排列起来，次序就固定不变了。多无聊的世界呀！

几千年过去了，直线国里生出一个叫卡鲁扎·克·莱茵（Kaluza K.Line）的，为压抑在直线上的人们带来一线希望。也许因为灵感，也许因为多年来看惯邻居的那“一点”眼睛而产生的幻想，总之，莱茵

猜测，直线国可能不是一维的。据他的理论，直线国实际上是二维的，第二维是卷缩着的小圆圈，因为在空间延展太小，所以还没有直接发现过它。他接着描绘了一种新的生命——假如那个卷缩的空间方向能够展开，那么照他的伙伴莱茵斯坦（Linestein）最近的研究，这种生命至少是可能的。莱茵描绘的世界令你和你的同伴们很兴奋，人人都满怀着希望——直线上的人们可以通过第二维自由地往来，受一维奴役的日子一去不复返了。我们看到，莱茵描绘的是一类生活在“有粗细的”水管世界的生命。

实际上，假如卷缩的小圆圈会长大，直线国“胀”成管子世界，你的生活也将发生巨变。以你的身体来说，在线形状态下，两眼间的一切构成你的身体。于是，对你来说，眼睛也就是皮肤，它将体内与体外的世界分隔开。直线国里的医生只有穿过眼睛才能给人做手术。

现在我们来看“胀大”的直线国会发生什么事情。我们假设卡鲁扎·克·莱茵理论中直线国的那一个隐藏卷缩的维展开来了，人人都能看到它。这时，别的线形生命能从侧面看到你的内部，见图8.5。通过展开的这一维，医生可以直接在暴露的身体内部动手术。这太不可思议了！看来，这些生命将“及时”长出一层皮肤来把暴露的内脏遮起来。而且，他们当然会进化成既有长度也有宽度的生命：在二维管子世界里滑行的平坦生命，如图8.6。假如卷缩的维足够大，这个二维宇宙就会像阿伯特的平直世界——一个假想的二维世界，有阿伯特赋予它的丰富的文化遗产，还有更具讽刺意味的以生命的几何形态为基础的社会等级。

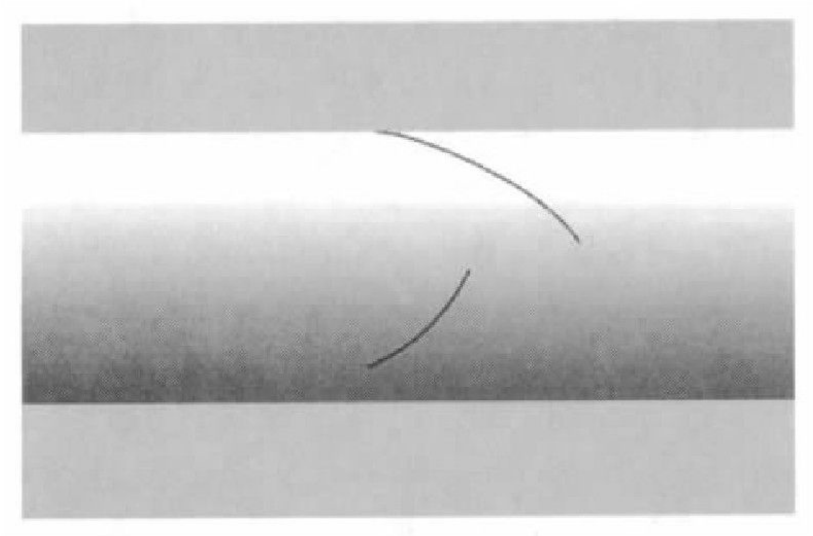


图8.5　直线世界膨胀为管子世界后，一个生命可以直接看到另一个生命的身体内部。

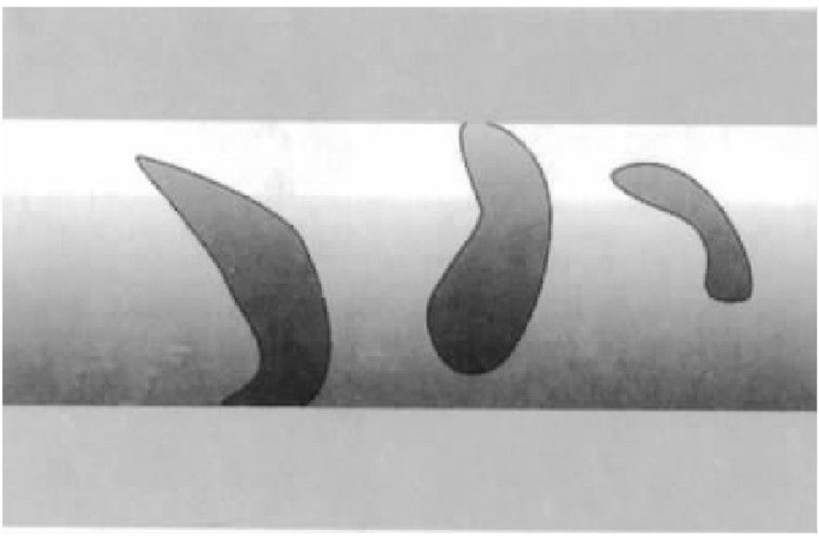


图8.6　生活在管子世界的平直二维生命。

在直线的世界里，我们很难想象能发生什么有趣的事情——因为没有足够的空间——在管子世界，好多事情都可能发生。从看得见的

一个大空间维进化到两个大空间维，真是“换了人间”。

现在，我们要问一个老问题：到此为止了吗？二维宇宙本身也可能有卷缩的一维，从而也可能是三维的。我们可以用图8.4来说明这一点，不过应该明白，我们现在想象的宇宙只有两个空间维（而在引进图8.4时，我们是用平面网格来代表3个展开的维）。如果卷缩的一维张开了，二维生命就会发现他生活在一个崭新的世界里，他不再限于两个方向的前后、左右运动了，现在，他也能在第三个方向——在那个圆圈维“上下”运动。实际上，如果这一维能长大，那就是我们的三维宇宙。我们现在还不知道我们的3个空间维是否会永远向外延伸，也许其中一维会卷缩成一个大圆，一个超出我们最大望远镜的大圆。假如图8.4的圆圈能长大——长到几十亿光年——那将是我们宇宙的良好写照。

不过，问题又来了：这就到头了吗？这将我们带近卡鲁扎和克莱茵的图景：我们的三维宇宙空间原本还有一个谁也不曾想到的卷缩的第四维。假如这惊人的图景——甚至更多维的更惊人的图景（我们很快会来讨论）——是真的，而且那些卷缩的维都展开来，成为宏观的维，那么根据刚才说的好几个低维的例子可以想象，我们的生命会发生多么大的变化。

令人惊讶的是，即使那些维总是小小的卷缩起来的，它们的存在仍然会产生深远的影响。

高维下的统一

我们宇宙的空间维数可能比我们直接感知的更多，卡鲁扎在1919年提出的这个建议从自身说来是很有可能的。不过，令它更动人的还在于别的原因。爱因斯坦在我们习惯的3个空间维和1个时间维的宇宙框架里建立了广义相对论，而这个理论的数学形式可以很直接地推广到更高维的宇宙，写下类似的方程。卡鲁扎在只多1个空间维的“最

保守的”假设条件下进行了这样的数学分析，具体导出了新的方程。

他发现，在修正了的形式中，与普通三维相关的方程从根本上说与爱因斯坦的方程是一样的。但是，因为他多包含了一个空间维，他当然也发现了爱因斯坦原来不曾导出的方程。在研究了这些与新维度相关联的方程后，卡鲁扎意识到有趣的事情正在发生。那多出的方程不是别的，正是麦克斯韦在19世纪80年代为描写电磁力而写下的方程！这样，通过添加1个空间维，卡鲁扎把爱因斯坦的引力理论与麦克斯韦的光的理论统一起来了。

在卡鲁扎的统一提出以前，引力和电磁力被认为是两种毫不相关的力，甚至没有一点儿线索暗示它们可能存在什么联系。卡鲁扎凭着他的创造力，大胆想象我们的宇宙还有另一个空间维，从而发现引力与电磁力实际上存在着深刻的联系。他的理论指出，两种力都伴随着空间结构的波动。引力在我们熟悉的3个空间维中波动，而电磁力则在那个新的卷缩的空间维里荡漾。

卡鲁扎把论文寄给爱因斯坦，爱因斯坦起初也很感兴趣。1919年4月21日，爱因斯坦回信告诉卡鲁扎，他从来没有想过统一能“通过一个五维[四维空间和一维时间]的柱形世界”来实现。他又补充说，“起初，我非常喜欢你的想法。”[44]可是，大约一个星期以后，爱因斯坦又来信了，这回他有点儿怀疑：“我读了你的文章，感觉它确实有意思。现在我还没有发现有什么不可能的地方。不过，另一方面，我得承认，目前提出的那些论证似乎还没有足够的说服力。”[45]两年多以后，爱因斯坦有了更多时间更彻底地消化卡鲁扎的新奇想法。1921年10月14日，他又写信告诉卡鲁扎，“再次觉得耽误了你发表你两年前关于引力和电力统一的思想……如果你愿意，我仍然可以把文章交给科学院。”[46]卡鲁扎终于收到了这位巨人迟到的“录取通知”。[47]

卡鲁扎的思想尽管很美妙，但后来经过克莱茵的仔细研究，发现它与实验结果有很大的矛盾。例如，一个简单例子是，把电子纳入理

论所预言的质量与电荷的关系，大大偏离了观测的数值。因为没有什么明显的办法来克服这个问题，许多关注卡鲁扎思想的物理学家也失去了兴趣。爱因斯坦等人还不时考虑过多余卷缩维的可能性，但它还是很快就离开了理论物理学的中心，成为一个边缘问题。

实在说来，卡鲁扎的思想走在了时代的前头。20世纪20年代标志着理论和实验物理学向微观世界的基本定律高歌猛进的开端。理论家们在全身心追寻量子力学和量子场论的结构；实验家们在忙着发现原子和无数其他基本物质构成的细节。理论指导实验，实验修正理论，这样经过半个世纪，物理学家终于找到了标准模型。在这果实累累令人振奋的年代里，多维的猜想当然只有远远躲到后面了。物理学家们在寻找有力的量子方法，寻找可以用实验来检验的预言，他们对多维空间的那点可能性不感兴趣——宇宙可能在小尺度下有迥然不同的面目，但那尺度却是我们最强大的仪器也无法探测的。

不过，激情的年代迟早会过去的。20世纪60年代末和70年代初，标准模型的理论结构成了新的潮流。到70年代末和80年代初，它的许多预言都被实验证实了，多数粒子物理学家相信，其他预言也终将被证实，那不过是时间问题。虽然好多具体问题还没有解决，但还是有很多人觉得，关于强力、弱力和电磁力的主要问题，已经有答案了。

最后我们又该回到那个最大的老问题：广义相对论与量子力学间的神秘的大冲突。三种力的量子理论已经成功建立起来了，这激励着物理学家们要把第四种力，引力，也囊括进来。他们尝试了数不清的方法，最终都失败了。所以，他们的思想也变得更加开放，也欢迎那些异乎寻常的思想方法。在20世纪20年代末被人遗忘的卡鲁扎-克莱茵理论，现在复活了。

现代卡鲁扎-克莱茵理论

自卡鲁扎理论提出60年以来，我们对物理学的认识发生了巨大的改变。量子力学完全确立了，也经过了实验的检验；20世纪20年代未知的强力和弱力也发现了，还有了深入的认识。有些物理学家提出，卡鲁扎最初的思想之所以失败，是因为他不知道那些其他的力，从而他对空间的变革还太保守。更多的力意味着需要更多的空间维。只凭一个卷缩的维——尽管能在广义相对论和电磁理论之间建立某种联系——还不足以结合更多的力。

20世纪70年代中叶，物理学家花了很大工夫来研究有多个卷缩空间方向的更高维理论。图8.7画了两个多余维的例子，那两维卷缩在一个球的表面，形成一个球面。跟一个卷缩维的情形一样，这些多余的维也生在我们熟悉的三维空间的每一点。（为清楚起见，我们只是在延展方向的网络点上画了二维的球面。）我们除了想象不同的维数，也可以想象多余的维有不同的形状。例如，图8.8画的也是两个卷缩维的一种可能情形，它们卷缩成面包圈的形状——也就是环。可以想象，还可能有更多的空间维，如3个、4个，5个甚至任意多个，可能卷缩成各种奇异的形状，可惜我们无法把它们画出来。这些维有一点是相同的：它们的空间延展都小于我们所能探测的最小尺度，因为我们还没有在实验中发现它们的存在。

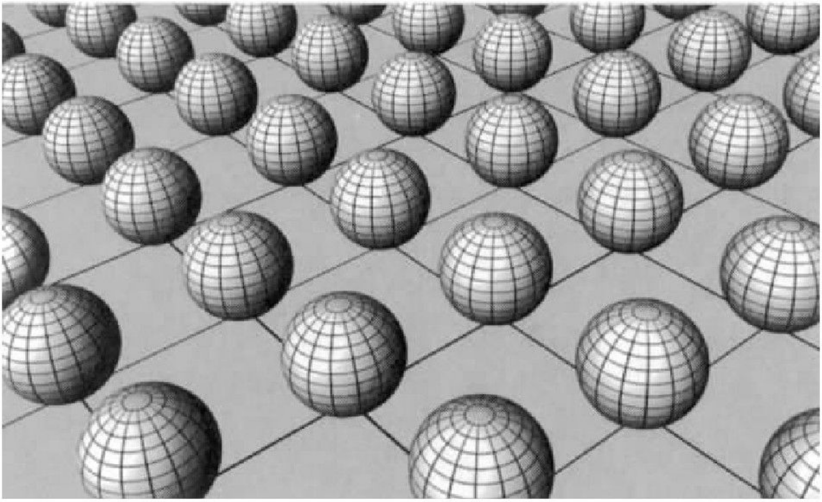


图8.7　卷缩成球面的两维。

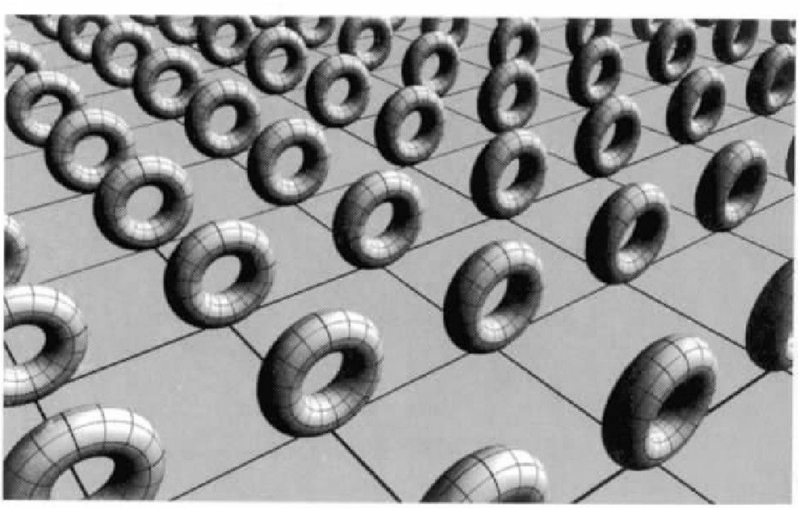


图8.8　卷缩成面包圈（环）的两维。

最有希望的高维想象是那些同时包含了超对称性的图景。超对称粒子对能部分消除许多剧烈的量子涨落，物理学家想靠它们来缓和广义相对论与量子力学间的矛盾。他们把这些包含引力、多维和超对称

性的理论称为高维超引力。

像卡鲁扎的原始想法一样，不同形式的高维超引力乍看起来似乎都有希望。从新维度产生的新方程会令人想起那些用来描写电磁力、强力和弱力的方程。不过，仔细考察会发现，老问题依然存在。最严重的是，令人讨厌的空间小尺度下的量子涨落虽然由于超对称性有所减弱，但还不足以产生一个合理的理论。物理学家还发现，很难找一个高维理论能把所有的力和物质特性都囊括进来。[3]

现在人们慢慢明白了，统一理论的碎片正在显现，但还缺少一条基本的线索把它们缝合起来成为一个与量子力学协调的大统一理论。1984年，那条失去的线索——弦——戏剧性地走进了我们的故事，站到了舞台的中心。

多维的弦理论

现在你该相信，我们宇宙可以包容更多的卷缩的空间维；当然，只要它们足够小，就没有东西能否定它们。但是，你也可以把多维当成一种技巧。我们看不见比百亿亿分之一米更小的距离，所以在那样的尺度下，不但多维是可能的，任何奇异的事情也都可能发生——甚至出现小绿人的微观文明。尽管多一些小空间维似乎比多一个小文明更合理，但不论设想什么，不经实验证明——在今天还不能证明——都同样是随意的。

弦理论出现以前的情形就是这样。我们需要一个理论来解决当代物理学面临的核心难题——量子力学与广义相对论的矛盾——并统一我们对自然基本物质组成和力的认识。但是，为了实现这些目标，弦理论要求宇宙有更多的空间维。

为什么呢？量子力学的一个主要观点是，我们的预言在根本上只能说某个事件会以某个概率发生。虽然爱因斯坦认为这是我们现代认

识的一个令人遗憾的特征，但你也可能看到了，那是事实，我们应该接受它。我们知道，概率总是0到1之间的数——当然，如果用百分数表示，也可以是0到100之间的数。物理学家发现，量子力学理论的某些计算得出的“概率”不在可以接受的范围，这是理论失败的信号。例如，我们在以前讲过，无穷大概率的出现，是点粒子框架下广义相对论与量子力学互不相容的信号。我们也讲过，弦理论能消除这些无限的东西；但我们没说还留着一个更玄妙的问题。在弦理论初期，物理学家曾发现某些计算会得出负概率，那也是不能接受的。这样看来，弦理论好像也淹没在它自己的量子力学的热浪里。

物理学家经过不懈努力，终于找到了负概率出现的原因。我们先来看一个简单的情形。假如一根弦束缚在二维面上——如桌面或者水管的表面，它就只能在两个独立方向振动：左右方向和前后方向。任何一个振动模式都是两个方向振动的组合。相应地，我们看到，在平直王国、管子世界或者其他二维宇宙的弦，也都只能在两个独立的空间方向振动。如果让弦离开二维面，那么它也能上下振动，这样独立的振动方向就增加到3个。就是说，在三维宇宙空间里，弦能在3个独立方向振动。依此类推（尽管难以想象），在更多空间维的宇宙中，弦能在更多的独立方向振动。

我们强调弦振动的事实，是因为物理学家发现那些令人困惑的计算结果强烈依赖于弦的独立振动方向的数目。负概率产生的原因就是理论需要的振动方向与实际表现的方向不相称：计算表明，如果弦能在9个独立空间方向振动，那么所有的负概率都将消失。这在理论上当然很漂亮，但那又如何呢？用弦理论来描写我们只有3个空间维的世界，我们似乎还是有麻烦。

真是那样的吗？半个多世纪过后，我们发现，卡鲁扎和克莱茵为我们留下一个窗口。因为弦很小，不但能在大的展开的空间方向振动，也能在小的卷缩的方向振动。这样，只要我们像卡鲁扎和克莱茵那样，假定在我们熟悉的3个展开的空间维以外还有6个卷缩的空间

维，就能在我们的宇宙中满足弦理论的9维空间的要求。弦理论就这样从物理学王国的边缘挽救回来了。而且，多维的存在，不仅是一种假定（如卡鲁扎、克莱茵和他们的追随者那样），更是弦理论的要求。为了让弦理论有意义，宇宙应该是10维的：9个空间维，1个时间维。这样，卡鲁扎1919的想象在今天找到了最有活力、也最有说服力的位置。

几个问题

这里生出几个问题。首先，为什么弦理论需要那样一个特别的空间维数来避免不合理的概率值呢？不借助数学公式，这大概是弦理论中最难回答的一个问题。直接用弦理论来计算能得到答案，但还没有人能用直观的非技术的方法来解释为什么会出现这个特别的数字。物理学家卢瑟福说过，大意是，如果我们不能以一种简单的非技术的方式解释一个结果，我们就还没有真正弄懂它。他不是说那个答案错了，而是说我们没有完全懂得它的起源、意义和作用。对弦理论的超维特征来说，这也许是对的。（顺便说一句，我们借这个机会来强调一下第12章将要讨论的第二次超弦革命的核心问题。关于十维时空——九维空间和一维时间——的计算后来证明是近似的。20世纪90年代中，惠藤根据他本人的发现和前人的一些结果（得克萨斯A&M大学的Michael Daff，剑桥大学的Chris Hull和Paul Townsend），提出了令人信服的证据，说明近似计算实际上丢失了一个空间维。他的结论令多数弦理论家大吃一惊：弦理论实际需要十一维，十维的空间和一维的时间。我们到第12章才讨论这个重要结论，现在忽略它不会给以下的讨论带来什么影响。）

第二，如果弦理论的方程（应该说近似方程；在第12章以前我们都在这个近似方程下讨论）证明宇宙有9个空间维和1个时间维，为什么其中的3个空间维（和那个时间维）是大的展开的维，而其余6个维是小的卷缩的呢？为什么它们不都展开或者卷缩？为什么不会是其他可能的情形呢？目前没人知道答案。如果弦理论是对的，我们总会找

出答案的，可我们对理论的认识，还不够深入，还回答不了这些问题。当然，这并不是说没人勇敢地尝试过回答它们。例如，从宇宙学的观点看，我们可以想象所有的维原来都是紧紧卷缩着的，然后，3个空间维和1个时间维在大爆炸中展开，一直膨胀到今天的尺度；而其余的空间维仍然卷缩在一起。至于为什么只展开了三维，我们也有大概的说法，将在第14章讨论。不过，实在说来，这些解释还只是略具雏形。在后面的讨论中，我们假定除了3个以外，别的空间维都是卷缩的，这是为了符合我们看到的周围世界。现代研究的一个基本目标就是确立这种假设来自理论本身。

第三，弦理论需要那么多额外的维，其中会不会有更多的时间维呢？那样不正好与多维的空间对应吗？用心想一想，你会发现那才真是令人困惑的事情。关于多维空间，我们总还有些认识，因为我们生活的世界一直都在与三维打交道。但多维时间意味着什么呢？难道一个时间跟我们寻常感觉和经历的时间相同，而另外的时间却多少有些“不同”？

当我们考虑卷缩的时间维，事情就更奇怪了。如果一只蚂蚁在卷缩成圆圈的空间爬行，爬过一圈，它总是回到原地。这一点儿也不奇怪，因为我们也总能回到空间的同一个地方，只要我们喜欢。可是，假如卷缩起来的是时间维，那么穿过它就意味着回去——在时间流过后回到以前的某一刻。这当然是我们没有经历过的。就我们的认识，时间是一维的，我们只能绝对地无选择地朝着一个方向走，永远也不可能回到它经过的瞬间。当然，卷缩的时间维在性质上也许不同于我们熟悉的那个从大爆炸创生长流到今天的大的时间维。但是，如果有新的以前未知的时间维，就不会像更多的空间维那么随意，虽然它们会更加“刻骨铭心”地改变我们对时间的感觉。有些理论家已经尝试过在弦理论中包容更多的时间维，但还没有什么结论性的东西。我们在讨论弦理论时，还是坚持更“传统的”观念，认为所有卷缩的维都是空间维。不过，在未来的理论中，新的时间维也许会扮演某个有趣的角

色。

多维的物理意义

从卡鲁扎的原始论文起，几十年的研究表明，尽管物理学家提出的额外的维都必须小于我们能直接“看到”的尺度（因为我们还没见过它们），但它们对我们看到的物理学确实有着重要的“间接的”影响。空间的这种微观性质与我们看到的物理学之间的联系在弦理论中表现得尤为显著。

为明白这一点，我们需要回想一下弦理论中的粒子质量和电荷是由可能的弦共振模式决定的。想象一根运动振荡的弦，你会发现它的共振模式受空间环境的影响。我们可以拿海洋的波浪来做例子。在无垠的大海，波可以相对自由地形成，以这样或那样的方式运动。这种情形很像振动的弦在大的展开的空间维度里穿行。我们在第6章讲过，这样的弦也可以在任何时刻在空间的任何方向自由振动。但是，假如海波经过狭窄的海湾，波形和运动肯定会受到水的深浅、岩石的形状和分布以及水道条件等因素的影响。当然，我们也可以想想单簧管或法国号，它们的声音是内部气流共振的结果，而这又取决于乐器中气流空间的形状和大小。卷缩的空间对弦的可能振动模式也会产生类似的影响。因为弦在所有空间维振动，所以那些额外的维如何卷缩、如何自我封闭，都强烈影响并束缚着弦的可能的共振模式。这些主要由额外维度的几何决定的模式构成了我们在寻常维度里可能观察到的粒子的性质。这就是说，额外维度的几何决定着我们在寻常三维展开空间里观察到的那些粒子的基本物理属性，如质量、电荷等。

这是极深刻而重要的一点认识，我们值得再说它一遍。照弦理论看，宇宙由一根根细小的弦构成，它们的共振模式就是粒子质量和力荷的微观起源。弦理论还要求所有多余的空间维都卷缩在极小的尺度里，难怪我们从来不曾见过它们。但是，小弦能探寻小空间。当弦振动着在空间运动时，多维的几何形态将决定它的共振模式。弦的共振

模式在我们看来就是基本粒子的质量和电荷，所以我们可以说，宇宙的这些基本性质在很大程度上决定于多余维度的几何形态和大小。这是弦理论的一个深远的洞察。

既然多余的维度那样深刻地影响着宇宙的基本物理性质，我们现在就带着无限的激情去看看那些卷缩的空间像什么样子。

卷缩的空间像什么

弦理论中的多余的空间维并不是随便能以任何方式“折皱”起来的；来自理论的方程严格限定了它们的形态。1984年，得克萨斯大学的坎德拉斯（Philip Candelas）、加利福尼亚大学的霍罗维茨（Gary Horowitz）和斯特罗明戈（Andrew Strominger）与惠藤证明，某类特殊的六维空间的几何形态能满足那些条件。那就是所谓的卡-丘空间（或卡-丘形态），是以宾夕法尼亚大学的数学家卡拉比（Eugenio Calabi）和哈佛大学的数学家丘成桐（Shing-Tung Yau）两人的名字命名的。他们两位在相关问题的研究比弦理论还早，对理解这些空间有着重要作用。尽管描写卡-丘空间的数学既复杂又玄妙，我们还是大概知道它们像什么样子。[4]

我们在图8.9画了一个卡-丘空间的例子。[48]你看这张图时，一定会感觉到它本来的局限——我们想在二维纸面上表现六维形态，当然会产生巨大的变形。不管怎么说，这图还是大致说明了卡-丘空间的样子。图8.9的形态不过是一个例子，还有成千上万的卡-丘形态都能满足在弦理论的额外维度所应具备的严格条件。虽然这种形态成千上万，似乎太多了，但与无限多的数字可能相比，卡-丘空间也实在是“稀有”的。

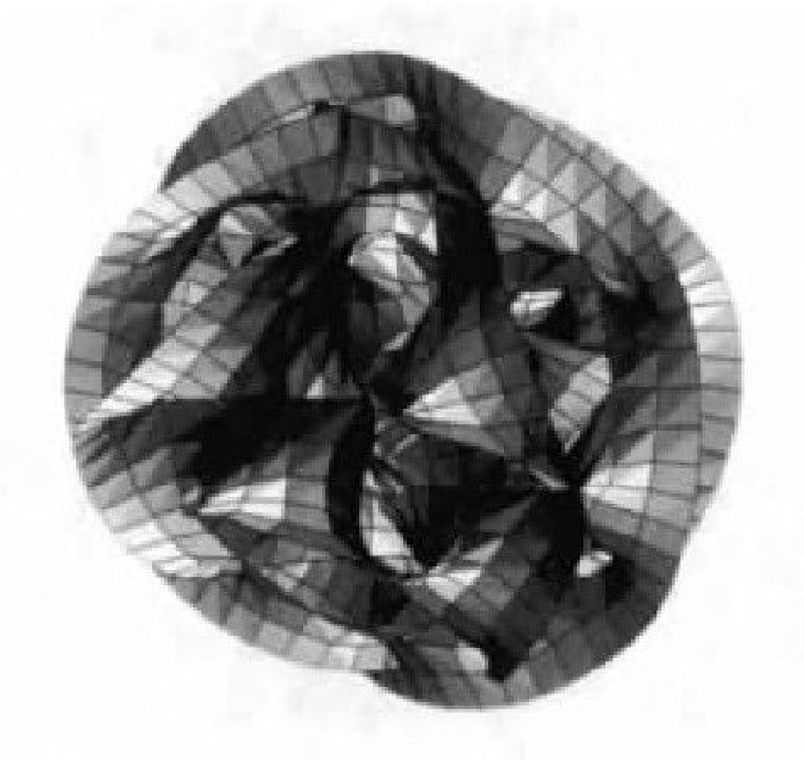


图8.9　卡拉比-丘成桐空间的一个例子。

好了，现在我们该用这些卡-丘空间来取代图8.7中代表两个卷缩维的球面。就是说，在寻常的三维展开空间的每一点生出一个弦理论所需要的六维空间，那些谁也不曾想过的维，紧紧地卷缩成一个看起来眼花缭乱的形状，如图8.10。这些维度无处不在，是空间结构不可分割的部分。假如你挥一挥手，你的手不但穿过三维展开的空间，也穿过了那些卷缩的空间。当然，卷缩的维太小，你的手不知扫过了多少那样的小空间。小空间的意思是没有大物体（如你的手）运动的余地——你的手挥过时，仿佛把小空间也“抹去”了，你根本不知道你自己经过了卷缩的卡-丘空间。



图8.10　根据弦理论，宇宙多余的维卷缩成卡拉比-丘成桐空间。

这是弦理论的一个惊人特征。但是，假如你想得更实际，你一定会把这些讨论与一个基本而具体的问题联系起来。既然我们对额外的维有了更好的认识，那么在这些空间振动的弦能生成哪些物理性质呢？这些性质又如何与实验观测相比较呢？那是弦理论中一个价值64000美元的问题。49

注释

1.这是一种简单的想法，但因为普通语言不够精确，常常引起误会，所以在这里澄清两点。第一，我们假定蚂蚁生活在管子的表面。如果蚂蚁钻进水管的内部——例如它穿透了橡皮水管——我们就得用3个数来确定它的位置，因为需要告诉它钻了多深。但如果蚂蚁只在水管表面活动，它的位置用两个数就能确定。这引出我们要讲的第二点：即使蚂蚁生活在水管表面，我们也可以（只要愿意）用3个数来确定它的位置：除普通的前后、左右方向外，还有它在我们熟悉的三维空间的上下方向的位置。但是，一旦我们知道蚂蚁只在水管的表面上，正文里说的两个数就够了，那是惟一确定蚂蚁位置的最少数据

——正因为这一点，我们说水管的表面是二维的。

2.令人惊奇的是，物理学家Savas Dimopoulos, Nima Arkani-Hamed和Gia Dvali在Ignatios Antoniadis和Joseph Lykken的研究基础上指出，即使多余的卷缩维有毫米大小，我们的实验仍然可能探测不到它们。原因是，粒子加速器是通过强力、弱力和电磁力来探测微观世界的。引力从我们技术能及的能量来说太微弱，一般是忽略了的。但Dimopoulos和他的伙伴们又指出，如果多余的维能对引力产生决定性影响（后来发现这在弦理论中是很可能的），则所有实验也都可能把它忽略了。在不远的将来，新的高灵敏引力实验会去寻找那样的“大”卷缩维。如果找到了，将是历史上最伟大的发现之一。

3.物理学家发现，高维理论最难应付的是标准模型的所谓手征性特征。为不使讨论过于沉重，我们在正文里没讲这个概念。但有些读者可能会感兴趣，所以在这里简单谈谈。假如有人让你看一段某个科学实验的影片，请你判断影片是实验本身的实况还是从镜子里看到的镜像。摄影水平很高，没留下镜子的一点儿痕迹。你能判断吗？20世纪50年代中，李政道和杨振宁的理论洞察，加上吴健雄和她的合作者们的实验，证明你能够做出判断，只要影片放的是某个适当的实验。换句话说，他们的研究表明宇宙不是完全镜像对称的——就是说，某些过程（那些直接依赖于弱力的过程）的镜像不可能在我们的宇宙发生，即使原过程可以发生。这样，如果你在影片中看到了不允许发生的过程，你就知道看的是实验的镜像，而不是实验本身。由于镜像交换左右方向，所以李、杨和吴的结果确定了宇宙不是完全左右对称的——用行话说，宇宙是具有手征性的。物理学家发现，正是标准模型的这一个特征（特别是弱力的），几乎不可能纳入高维的超引力框架。为避免混淆，这里说明一点，我们在第10章将讨论弦理论的“镜像对称”概念，那里的“镜像”与这里讲的是完全不同的。

4.懂数学的读者应该知道，卡拉比-丘成桐流形是第一陈（省身）类为零的一种复Kähler流形。1957年，卡拉比猜想所有这类流形都存

在平坦的Ricci度规，1977年，丘成桐证明猜想是正确的。

## 第9章　证据：实验信号

最令弦理论家高兴的，莫过于堂堂正正地向世界提出一系列详细的能经受实验检验的预言。当然啦，没有经过实验验证的理论是不可能用来描写世界的。不论弦理论描绘的图景多诱人，如果它没能准确描写我们的宇宙，就不过是精巧的“地下城与龙”的游戏。

惠藤骄傲地宣布，弦理论已经做出了激动人心而且经实验证实了的预言：“弦理论具有一个令人瞩目的性质，它预言了引力。”[50]惠藤这话的意思是，牛顿和爱因斯坦都是因为他们对世界的观察表明存在着引力需要一个准确而和谐的解释，才去创立他们的引力论；而另一方面，研究弦理论的物理学家，即使一点儿不懂广义相对论，也会不可避免地在弦的引导下走向它。弦理论通过零质量的自旋-2引力子振动模式，把引力密密地织入了它的理论结构。正如惠藤说的，“弦理论产生引力论，这是空前伟大的理论发现。”[51]惠藤也承认，所谓的“预言”应该再贴上“后言”的标签，因为物理学家早在知道弦理论前就发现了引力的理论描述。但他又指出，这不过是地球上的历史巧合罢了。他猜想，在宇宙的其他高等文明里，很可能先发现弦理论，后来才发现引力理论是它的一个动人结果。

我们当然尊重自己星球上的科学史，有很多人认为，所谓引力的“后言”并不是弦理论令人信服的实验证明。多数物理学家更喜欢真正的预言或“后言”，它们要么能通过实验来证明，要么是目前还不能解释的宇宙的某些性质（如电子的质量，或存在三族粒子等）。我们这一章将讨论弦理论家朝着这个目标走了多远。

有讽刺意味的是，我们将看到，尽管弦理论可能是物理学家研究过的最具预言能力的理论——一个有能力解释最基本的自然性质的理论——但物理学家却还拿不出一个足够精确的能面对实验数据的预

言。小孩得到了梦想的圣诞礼物，却不知道怎么玩儿——说明书丢了几页。今天的弦理论家也处在这种境地，他们手里可能正握着现代物理学的圣杯，却发挥不了它预言的威力，因为完整的使用手册还没写好。不管怎么说，我们要在这一章说明，运气好的话，弦理论的一个核心特征在10年后可能得到实验验证。如果运气更好些，我们随时都可能证实理论的一些间接特性。

四面楚歌

弦理论对吗？我们不知道。如果你也相信物理学定律不该分离成大的和小的两个领域，而且还相信我们应该永不停息地寻找一个没有应用极限的理论，那么，弦理论就是惟一值得考虑的途径。不过，你可能认为，那只能说明物理学家缺乏想象，并不说明弦理论有什么基本的惟一性。也许真是这样。你也可能会说，物理学家留连于弦理论，只是因为变幻多姿的科学史恰好在这个方向上投来一丝光亮，这就像丢了钥匙的人只在街灯的昏暗光影里去寻找。这也可能是真的。而且，假如你有些保守，或者爱玩儿些诡辩，你甚至还可能说，物理学家无权把时间浪费在这样一个幻想的理论上，它所提出的那些自然新特征比我们实验直接探测的任何事物还小几乎20个数量级。

如果你是在20世纪80年代弦理论刚闪亮登场时发这些抱怨，可能我们今天的大多数物理学家都会有同感。例如，80年代中期，哈佛大学的诺贝尔奖获得者格拉肖，还有物理学家金斯帕格（Paul Ginsparg，那时也在哈佛），曾公开批评弦理论没有实验检验的可能：

超弦理论追求的不是传统的理论与实验的统一，而是一种内在的和谐，以精密、独特和优美来决定真理。这个理论的存在，靠的是一些魔幻般的巧合，无穷大在这里奇迹般地消失了，看似毫无关联（也可能尚未发现）的数学领域也奇迹般地联系起来了。难道这些性质能成为我们把超弦当成实在的理由吗？难道数学和美学就这样完全替代

并超越实验了吗？[52]

在别的场合，格拉肖又说，

超弦理论野心勃勃，它要么完全正确，要么完全错误。惟一的问题是数学太新、太难，再过几十年我们也不知道会出现什么结果。[53]

他甚至提出“物理系是否还应该为弦理论家们掏钱？还让他们去误导不懂事的学生吗？”他警告大家，弦理论在损害着科学，跟中世纪的神学没什么两样。[54]

费曼在去世前明确表示，他不相信弦理论是解决困扰引力与量子力学和谐统一的问题——特别是令人讨厌的无限大问题——的惟一良方：

我的感觉是——可能是错的——解决问题的途径有很多。我想不会只有一种办法才能摆脱那些无穷大。对我来说，一个理论只凭摆脱了无穷大，还不足以令人相信它就是独一无二的。[55]

格拉肖在哈佛的同事和伙伴乔基，在20世纪80年代末也是弦理论的积极批评者：

假如我们甘愿沉溺于那种在我们的实验朋友无能为力的小距离尺度的“终极”统一的诱惑，我们就会陷入困境，因为那样我们将无法剔除那些不相干的东西，而正是这个过程才使物理学不同于许多别的不那么有趣的人类活动。[56]

同许许多多的大问题一样，有积极的反对者，也会有热情的支持者。惠藤说过，当他知道弦理论如何把引力和量子力学结合在一起时，他经历了有生以来“最强烈的思想震撼”。[57]著名弦理论家、哈佛大学的瓦法（Cumrun Vafa）说，“弦理论无疑前所未有地揭示了宇宙最深层的东西。”[58]诺贝尔奖获得者盖尔曼也说，弦理论是“很迷人的

东西”，他盼着它的某种形式能在某一天成为整个世界的理论。[59]

我们看到，论战发生在物理学和关于物理学该怎么做的形形色色的哲学之间。“传统论者”希望理论工作走几百年来的成功之路，紧紧与实验观测相联系。但另一些人则认为我们有能力解决当今实验技术不能直接检验的问题。

尽管众说纷纭，在过去的十年间，对弦理论的批判慢慢平息了。格拉肖认为有两个原因。第一，在20世纪80年代中期，弦理论家们曾狂热而野心勃勃地宣扬他们将很快回答物理学的所有问题。现在，他们谨慎多了，我在20世纪80年代的许多批评没有意义了。[60]

第二，他又指出

我们这些不是弦理论家的人在最近十年里什么进展也没有，关于所以，弦理论是惟一途径的说法，是很有力量的。许多问题不能在传统量子场论的框架下解决，这是明摆着的事情，它们可能由别的东西来解决，而据我所知，那东西就是弦理论。[61]

乔基差不多也是这样回顾20世纪80年代的：

弦理论在发展之初的许多时候被宣扬过头了。这些年里，我发现弦理论的某些思想引出了有趣的物理学思路，对我自己的研究也很有帮助。现在我更高兴地看到人们在弦理论上付出辛劳，因为我能明白那些有用的东西将如何从中产生出来。[1]

格罗斯既是传统物理学家，也是弦理论家，他生动地总结了弦理论的状况：

我们像是在攀登大自然这座山，实验家总是赶在前头，我们这些懒散的理论家老是落后。他们偶尔踢下一块石头，砸在我们头上。最终我们会觉悟，并沿着实验家们开辟的路往前走。当我们与实验家走

到一起时，我们会告诉他们，我们觉悟了什么，是如何觉悟的。这是最传统也最容易的（至少对理论家来说）登山途径。我们都向往着能回到那些日子。但是现在，我们理论家可能不得不赶到前头了，这是更加孤独的征程。[62]

理论家并不想在自然的山峦独自登高，他们更愿与实验伙伴们共同经历艰辛，分享快乐。可惜的是，我们的历史不同步，今天的状况不够和谐，理论的登峰工具齐备了，实验的还没有。但这并不是说弦理论与实验分道扬镳了。实际上，弦理论家很可能“踢下一块理论的石头”，从超高能的山巅滚落到山下的实验家们的大本营。这是当今弦理论研究的基本目标。当然，还没有哪块石头从山巅飞落下来，但正如我们现在讲的，的确有几块诱人的石头正摇摇欲坠呢。

走向实验

如果没有大的技术突破，我们永远也不可能聚焦到能直接看到一根根弦的小尺度上来。物理学家可以用几英里大的加速器探测100亿亿分之一米大小的尺度。探测更小的尺度需要更高的能量，这意味着把能量聚集到单个粒子的机器也应该更大。由于普朗克长度比我们今天能达到的最小尺度低17个量级，用今天的技术，要银河系那么大的加速器才能直接看见一根一根的弦。实际上，特拉维夫大学的努辛诺夫（Shmuel　Nussinov）已经证明，这个基本的直观尺度的粗略估计似乎太乐观了，他更详细的研究表明，我们需要的加速器该有整个宇宙那么大。（探测普朗克长度下的物质所要求的能量大约等于1千千瓦小时，差不多是普通空调工作1千小时的耗电量——这看来也不怎么稀奇。最大的技术难题在于如何把这个能量完全集中到一个基本粒子，即一根弦上。）美国国会最终取消了超导超级对撞机（SCS）的资助——那“不过”才86千米的周长——所以，我们用不着焦急盼望有人会拿钱来做普朗克的加速器。如果我们还想用实验来检验弦理论，那只能用间接的方法。我们只好找出弦理论的某些物理结果，在比弦本身尺度大得多的尺度下去观测它们。[2]

坎德拉斯、霍罗维茨、斯特罗明戈和惠藤在他们“破土奠基”的文章里，向着这一目标迈出了第一步。他们不但发现弦理论中多余的维度应该卷缩成卡-丘空间的形态，还计算了一些可能对弦振动模式产生影响的结果。他们发现的一个主要结果表明，弦理论可能为存在已久的粒子物理学问题带来了令人意想不到的答案。

回想一下，物理学家发现的基本粒子分成三个组织相同的族，后一组比前一组有更大的质量。弦理论出现以前，有一个问题一直令人困惑：为什么粒子成族出现？为什么是三族？弦理论是这样考虑的：典型的卡-丘空间都包含着洞，像唱片或面包圈，甚至像“面包圈链”，如图9.1。在高维卡-丘空间背景下，实际上有多种不同类型的孔——孔本身可以有不同的维（“多维孔”），但图9.1说明了基本思想。坎德拉斯等人认真考察了这些孔对弦振动模式可能产生的影响，下面是他们的发现。

空间的卡-丘部分的每一个孔都关联着一组最低能量的弦振动模式。因为我们熟悉的基本粒子都该对应于最低能量的振动模式，所以，多孔的存在（像多孔的面包圈）意味着弦振动模式应该是多族的。假如卷缩的卡-丘空间有3个孔，那我们就会看到三族基本粒子。[3]这样，弦理论告诉我们，实验观察到的粒子族组织，不是什么解释不了的源于随机或神奇的特征，而是构成多维空间的几何形态的孔数的反映！这类结果令物理学家心动不已。

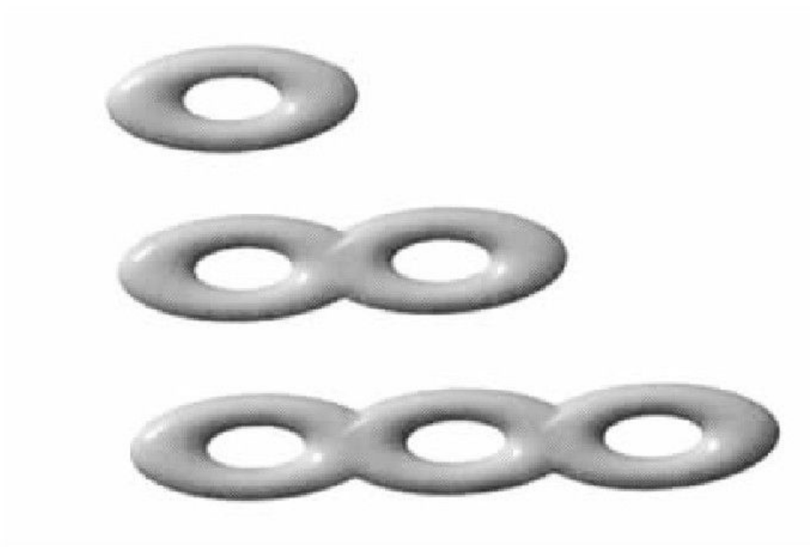


图9.1　面包圈（环）和它的多孔伙伴。

也许，你认为卷缩的普朗克尺度的空间维的孔数——卓绝的“山顶物理学”——就是一块落到一般能量下的试金石。毕竟，实验家能够——实际上已经——确定三族粒子与那孔数相对应。遗憾的是，已知的成千上万的卡-丘空间包含的孔数各不相同，有的是3，但也有4，5，25的，甚至还有多达480的。现在的问题是，没人知道如何从弦理论方程导出哪些卡-丘形态构成了额外的空间维。假如我们能找到一个能从无数可能中挑选出某个卡-丘形态的原则，那么，石头就真的从山巅滚落到实验家的大本营来了。假如从方程中选出的特殊卡-丘形态一定有3个孔，我们便从弦理论发现了动人的“后言”，解释了本是一团迷雾的已知的自然特征。但我们现在还没有发现那样的选择原则。不管怎么说——这也是很重要的——我们看到弦理论具有回答粒子物理学基本疑难的潜力，这本身就是一大进步。

粒子的族数不过是多维几何形态的一个实验结果。通过对弦振动模式产生影响，多维的结果还包括力和物质粒子的具体性质。看一个基本例子：斯特罗明戈和惠藤后来发现，每一组粒子的质量依赖于

——或者说取决于——卡-丘空间中各种多维孔洞边界是如何交叉和重叠的。这个问题有点儿复杂，很难形象表达。大概意思是说，当弦在卷缩维振动时，卡-丘空间孔洞的分布和孔洞周围的空间褶皱方式将直接影响可能的共振模式。细节很难讲，也并不都很重要；重要的是，跟粒子族的情形一样，弦理论还能提供一个框架来回答为什么电子和其他粒子的质量是那样的，等等诸如此类的问题——以前的理论对这些问题无话可说。不过，完成这些计算还是需要我们知道多余的维具有哪样的卡-丘空间形态。

上面的讨论大概说明了，弦理论如何可能在未来的某一天解释表1.1列举的物质粒子的性质。弦理论家相信，根据同样的理由，它还可能解释表1.2列举的基本力的信使粒子的性质。就是说，当弦在展开和卷缩的空间里卷曲振动着运动时，无数振动模式中的一小部分构成自旋等于1或2的集合，这些可能就是传递力的弦振动状态。不管卡-丘空间是什么形态，总会有一种质量为0、自旋为2的振动模式，我们说它就是引力子。不过，自旋为1的信使粒子——它们的数目，所传递力的强度以及它们遵从的规范对称性——则强烈依赖于卷缩维的具体几何形态，我们还不能完全列举出来。这样，我们又一次看到，弦理论提供了一个框架，能解释我们观察到的宇宙的信使粒子的性质，也就是能解释基本力的性质。但是，我们还不知道那些多余的维卷缩成了哪种卡-丘空间形式，所以还得不出确定的预言或“后言”（除了惠藤讲的关于引力子的后言而外）。

我们为什么选不出那个“正确的”卡-丘空间形态呢？多数弦理论家将它归咎于我们今天用来分析弦理论的工具还不够充分。我们在第12章会更详细地讨论，弦理论的数学工具太复杂了，物理学家只能在所谓微扰论的形式下做一些近似计算。在这近似的框架下，所有卡-丘空间似乎都是平等的，方程决定不出一个基本的形态。由于弦理论的物理结果敏感地依赖于卷缩维的准确形态，不能从大量卡-丘形态中选出一个，就不可能得到确定的能用实验检验的结果。今天研究背后的一

大动力就是发展超越近似方法的理论方法，希望它能带来一些结果，特别是将我们引向一个惟一的多维的卡-丘空间形态。我们将在第13章讨论这些路线取得的进展。

数不尽的可能

于是你可能要问：即使我们还不知道弦理论选择的是哪种卡-丘形态，那么，任选一种形态能得出与我们的观测一致的物理性质吗？换句话讲，假如我们把与每一种卡-丘形态相关联的物理性质都找出来，然后汇集在一起，我们能找出与实在相符某个形态吗？这是一个很重要的问题，但主要因为两点理由，我们很难完全回答它。

我们先来看产生3族的卡-丘形态，这应该是合理的出发点。它大大削减了可能的选择，但还是有很多。实际上，我们可以让面包圈变形，从一个形态变成许多形态——其实是无穷多——而不会改变孔的数目。在图9.2中，我们将图9.1下面的三孔圈变成现在这样。同样，我们可以从一个三孔的卡-丘空间开始，光滑地改变它的形态而不改变孔数，这样又生成一个无限的形态序列。（我们以前说万种卡-丘形态，已经把能相互光滑变形的空间合并成一组，这样的一组算一种空间形态。）问题是，弦振动的具体物理性质（它们的质量、它们对力的响应）严格受空间具体形态改变的影响，而我们仍然没有办法选择最可能的形态。不论教授让多少研究生去做，也不可能列出对应于无穷多空间形态的物理学。

认识到这一点后，弦理论家便去考察从可能的卡-丘形态的某些样本能生成什么物理学。然而，即使在这种情形，他们也不是一帆风顺的。理论家们现在用的近似方法并不是从一定的卡-丘形态导出所有的物理学。从粗略的意义说，它们会大大有助于我们理解那些我们希望能与观察到的粒子相对应的弦振动；但是，为得到精确确定的物理学结果（如电子的质量、弱力的强度），我们需要比今天的近似框架精确得多的方程。想想我们在第6章讲的，弦理论的“自然”能量尺度是

普朗克能量，只有经过“价格游戏”那样极端精巧的能量消减，才能得到具有已知物质和力的粒子质量的弦振动模式。精巧的能量消减靠的是精确的计算，哪怕是很小的误差，也会对精度产生巨大的影响。正如我们将在第12章讨论的，物理学家20世纪90年代中期在超越目前近似方程上已经取得了重大进展，当然，前面的路依然很长。

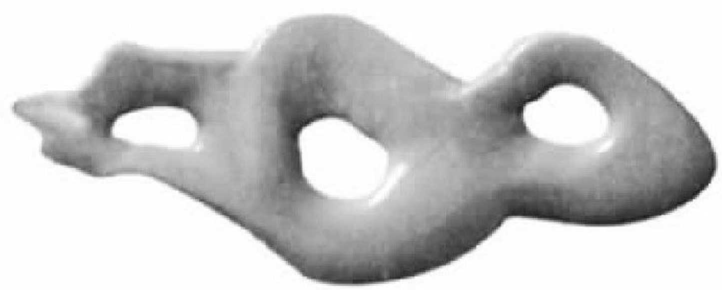


图9.2　多孔面包圈可以通过多种方式变形，而不会改变孔的数目。这是一个例子。

那么，我们现在的情形怎样呢？虽然没有一个基本准则指导我们选择一个卡-丘空间形态，也没有足够的理论工具从那样的选择中得出所有的可观测结果，但我们还是可以问，是不是任何一个卡-丘形态的选择都能产生一个与我们的观察一致（哪怕是大体一致）的世界？答案是令人鼓舞的。尽管多数卡-丘空间生成的结果与我们的世界迥然不同（不同数目的粒子族，不同数目和类型的基本力，以及其他许多不同的东西），但还是有几种选择的物理学在性质上确实与我们实际看到的相同。那就是，有些卡-丘空间在选择为弦理论所要求的卷缩维的形态时，产生的弦振动非常接近标准模型的粒子。而且，特别重要的是，弦理论成功地将引力编织进了量子力学的框架。

就我们现在的认识水平，这样的局面已经够好了。假如很多卡-丘形态都能与实验大体相符，特别的某个选择与我们观察的物理学之间

的联系就不那么令人感兴趣了。许多选择都能满足，那么即使从实验观点看似乎也选不出一个特别的来。另一方面，假如没有一个卡-丘形态能产生我们看到的物理学性质，那么弦理论就与我们的世界无关，虽然它的理论结构是那样美妙。我们今天决定具体物理学结果的本领还低得可怜，凭这点能力，找少数几个卡-丘形态，能在粗略水平上令人接受，就是很令人鼓舞的结果了。

解释基本物质和力的粒子性质，至少应该是最伟大的科学成就之一。不过，你可能还是要问问，不论现在或是不远的将来，会不会有什么真正的弦理论的预言——不是“后言”——能让实验物理学家来证实？是的。

超粒子

从弦理论导出具体的预言，眼前还有许多理论障碍，这迫使我们去寻找由弦构成的宇宙的一般而不是特殊的方面。这里的一般，说的是弦理论的那样一些基本特征，它们几乎（如果不是完全的话）不受超出我们现在理论水平的那些具体性质的影响。即使我们不懂得整个理论，还是可以满怀信心地讨论这些一般特征。在以后的篇章里我们会讲很多例子；现在我们先看一点：超对称性。

我们曾经讲过，弦理论的一大基本特征是它具有高度的对称性，它不仅包含了直观的对称性原理，还遵从这些原理的最大的数学扩张——超对称性。正如第7章讲的，这意味着弦振动模式是成对产生的——所谓的超对称伙伴对——一对伙伴的差别仅在于差半个自旋单位。如果弦理论是正确的，那么某些弦振动将对应于已知的基本粒子，由于超对称伙伴的出现，弦理论也预言每个这样的基本粒子都应该有一个超对称伙伴粒子。我们可以确定这些超伙伴粒子该携带多大的力荷，却还没有办法预言它们的质量。即便如此，超伙伴存在的预言是弦理论的一般特征之一；它是真正的弦理论的性质，与我们尚未明白的理论的其他方面无关。

然而，我们从没发现过已知粒子的超对称伙伴，这似乎说明它们并不存在而弦理论错了。不过，许多粒子物理学家认为，那说明超伙伴太重了，超出了我们今天的实验观测能力。现在，物理学家还在瑞士日内瓦做庞大的加速器，叫大型重子对撞机。这台机器很有希望发现超伙伴粒子。它在2010年以前大概就能运行了，不久超对称性就可得到实验证明。正像施瓦兹说过的，“发现超对称应该不会等得太久；那一天的到来一定是激动人心的。”[63]

不过，我们应该牢记两件事情。即使超伙伴找到了，仅凭这一点也不能保证弦理论是正确的。正如我们看到的，尽管超对称是在弦理论研究中发现的，但它也成功走进了点粒子理论，从而并不惟一属于弦。反过来讲，即使大型重子对撞机发现不了超伙伴粒子，这一点也不能排除弦理论，因为超伙伴的质量也可能超过了那台机器的能力。

话虽这样说，假如超伙伴真的发现了，对弦理论来讲肯定是一个强有力的令人振奋的间接证据。

分数电荷

弦理论的另一个实验信号与电荷有关，似乎不像超伙伴粒子那么“一般”，但也同样激动人心。标准模型的基本粒子的电荷只有有限的几种：夸克和反夸克的电荷是1/3、2/3和-1/3、-2/3；其他粒子的电荷为1，0和-1。这些粒子的组合能解释宇宙间所有已知的物质。然而，在弦理论中，可能存在一些共振模式对应着电荷大不相同的粒子。某些粒子可能具有非常奇怪的分数电荷，如1/5、1/11、1/13或1/53等。这些异乎寻常的电荷可以来自一定几何性质的卷缩维：空间的孔有那种特殊性质，绕着它们的弦需要绕过一定的圈数才可能自行解开。[4]细节并不重要，重要的是圈的数目在可能的弦振动模式中表现出来了，那就是分数电荷的分母。

有些卡-丘空间有这种几何性质，而另一些没有，因此分数电荷的

可能出现并不像超伙伴粒子的存在那样“一般”。另一方面，超伙伴的预言不是弦理论的独特预言，而几十年的经验告诉我们，任何点粒子理论似乎都没有充分理由存在这些奇异的分数电荷。如果谁要硬把这些电荷塞进点粒子理论，他可就成了瓷器店里横冲直撞的大公牛。分数电荷从额外空间维可能具有的简单几何性质突现出来使这些奇异的电荷成了检验弦理论的自然的实验信号。

跟超伙伴的情形一样，那种带奇异分数电荷的粒子我们也从没见过，而我们对弦理论的认识也还不能确定地预言它们的质量——假如卷缩的维真有产生它们的恰当性质的话。看不到它们的原因还是那句老话：如果确实存在，它们的质量一定超出了我们目前的技术能力——事实上，它们的质量可能是普朗克质量级的。但是，假如未来某个实验遇到了这种奇异的粒子，那将成为弦理论的一大证据。

几点猜想

我们还可能通过别的方法找到弦理论的证据。例如，惠藤曾提出一个大胆的猜想；天文学家可能有一天会在他们收集的天文数据里发现直接的弦理论信息。我们在第6章说过，弦的典型尺度是普朗克长度，但高能的弦可以大得多。实际上，大爆炸的能量可能足以产生几根从宏观上看也足够大的弦，这些弦随着宇宙膨胀可能长到天文学的尺度。我们可以想象，一根这样的弦可能在现在或者将来的某一天扫过夜空，在天文学家们收集的数据里留下醒目而可测的印迹（如微波背景辐射的温度出现小小偏移；见第14章）。正如惠藤说的，“尽管那多少是个幻想，但却是我最欣赏的证实弦理论的图像，因为没有什么能比在望远镜里看到一根弦更激动人心的事了。”[64]

距离地球近一些的可能的弦理论实验信号也有人提出来了。我们看五个例子。第一，我们在表1.1中说过，我们不知道中微子到底是质量很小，还是根本没有质量。根据标准模型，它们是没有质量的，但并没有什么特别深刻的原因。弦理论面临的一个挑战就是，为现在

或将来的中微子数据找一个令人信服的解释——特别是，如果实验最终证明它确实具有小小的非零质量。第二，某些标准模型禁戒的假想过程在弦理论中却是可能发生的。例如，质子可能分解（别担心，即使真有这种分解，也是十分缓慢的），不同夸克的组合可能相互转变或者衰变，这些都违背了点粒子量子场论中的某些确立已久的性质。[5]这些过程之所以特别有意思，是因为它们是传统理论没有的东西，从而也成为新物理的一个敏感信号：如果不求助新的原理，就解释不了它们。如果能观测到这些过程发生，那么任何一个都能为生成弦理论的解释提供肥沃的土壤。第三，某些卡-丘空间形态的选择会出现特别的弦振动模式，对应于一些新的小的长程作用的力场。假如这些新力的效应发现了，它们可能反映弦理论的某些物理特征。第四，正如我们将在下一章看到的，天文学家收集了大量证据说明我们银河系甚至整个宇宙都浸没在所谓暗物质的汪洋里，但暗物质至今还没得到确认。弦理论通过多种可能的弦振动模式，提供了许多暗物质候选者；等将来实验结果揭示出暗物质的具体性质以后，我们才能确定某个候选者。

最后，联系弦理论与实验观测的第五种可能途径牵涉到宇宙学常数——我们还记得，在第3章讨论过的，这是爱因斯坦为了保证一个静态宇宙而临时在他原始的广义相对论方程里添加的一个修正参数。后来发现宇宙在膨胀，爱因斯坦便取消了这一项，但物理学家从那时就认识到，我们不能解释为什么宇宙学常数应该是零。实际上，宇宙学常数可以解释为某种存在于真空的能量，从而应该可以根据理论计算它的值，也可以用实验来测量。但在今天，那些计算与观测却带来一大堆的矛盾：观测表明，宇宙学常数要么为零（如爱因斯坦最终认为的），要么很小。计算表明，虚空的量子力学涨落可能生成一个非零的宇宙学常数，比实验允许的大120个数量级（1后面跟120个零）！这向弦理论提出了一个挑战，也提供了一次机遇：弦理论的计算能与实验对应起来吗？能解释宇宙学常数为什么是零吗？或者，假如实验最终确定它的值很小但不是零，弦理论还能解释吗？假如弦理

论家能响应挑战——现在还没有呢——将为理论带来多么激动人心的支持啊！

评判

物理学史充满了那样的思想，它们在刚提出时似乎完全不可能证实，但经过意想不到的发展以后，最终还是走进了实验的王国。原子的思想、泡利的中微子假设，中子星和黑洞的预言，都是这样的例子——这些东西，我们现在完全相信了，当初它们却更像科幻小说的玄想，没有一点儿科学事实的影子。

弦理论出现的原因，至少跟那三个例子一样动人——事实上，我们曾经欢呼，弦理论是自量子力学发现以来最重要最激动人心的理论物理学进步。拿这两者来比较是很恰当的，因为量子力学的历史告诉我们，物理学革命有时要经历几十年才能走向成熟。与今天的弦理论相比，量子力学战线的物理学家该是很幸运的：量子力学在尚未完全建立的时候，就能直接与实验结果发生联系。即使这样，量子力学的逻辑结构也过了近30年才建立起来，而又过了20年才完全与狭义相对论结合在一起。我们现在要把它与广义相对论结合起来，是更富挑战性的使命；而且，与实验相联系更是难上加难。与量子理论的开拓者们不同的是，弦理论家没有看到一丝自然的光亮透过具体的实验结果来指引他们一步步往前走。

这就是说，弦理论的认识和发展可能在耗尽一代或几代物理学家的心血后，还得不到一点儿实验响应。世界上热烈追求弦理论的数不胜数的物理学家都知道，他们是在冒险：一生的奋斗可能只换来飘忽不定的结果。当然，理论总会进步的，但它能克服今天的障碍而得到确定的能让实验检验的预言吗？我们上面讨论的间接检验能为弦理论带来确凿的证据吗？这些问题每一个弦理论家都很关心，但谁也说不清一点儿东西。我们只有等着答案的到来。美妙而简洁的形式，强大的囊括万物的力量，无限的预言能力，简单而自然的消除引力和量子

力学矛盾的方式，弦理论的所有这一切，荡起无数人的激情，甘愿为它冒巨大的风险。

这些崇高的愿望在一点点地变得更实在——弦理论不断揭示出弦宇宙的新物理学特征——那些揭示了大自然杰作中更微妙更深层联系的特征。用上面的话讲，多数这样的特征都是一般性的，不论我们今天未知的东西怎样，它们都是弦构成的宇宙的基本特征。在这些特征里，最惊人的那些已经对我们不断演进的时空认识产生了深远的影响。

注释

1.这是1997年12月28日我访问乔基时他对我说的。这次访问中，乔基还告诉我，当实验否定了他和格拉肖在大统一理论中最先提出的质子衰变的预言时（见第7章），他对超弦理论感到犹豫不决。他尖锐指出，他的大统一理论所借助的能量比以前任何理论所考虑的都高得多；而当预言被证明是错误的时候——当他“被大自然压垮”的时候——他研究高能物理学的态度忽然改变了。我问他，如果实验证明了他的大统一理论，会激发他去关心普朗克尺度吗？他回答说：“是的，我很可能会的。”

2.说到这里，应该记住第6章后面注释3提出的那个猜想，弦只是可能比原来想的长得多，从而有可能在几十年内通过加速器来接受实验的检验。

3.对数学感兴趣的读者应该看到，更准确的数学表述是，粒子族的数目是卡-丘空间欧拉数的绝对值的一半。欧拉数本身是流形的同调群维数的交错和——在这里我们粗略地把同调群当作多维的孔洞。这样，从欧拉数为±6的卡-丘空间生成3个粒子族。

4.对数学感兴趣的读者知道，我们这里说的是具有有限非平凡基本群的卡-丘空间，群的阶数在某些情况下决定了分数电荷的分母。

5.专业读者知道，这里有些过程破坏了轻子数守恒定律和电荷-宇称-时间（CPT）反演对称性。

# 第四篇
# 弦理论与时空结构

## 第10章　量子几何

在大约10年的时间里，爱因斯坦凭他的一双手推倒了200多年老的牛顿体系，为世界带来了可以证明的崭新而深刻的引力理论。不论专家还是外行，都喜欢谈爱因斯坦在塑造广义相对论时所表现的卓绝才华和惊人的创造力，不过，我们也不应该忘记对他的成功有过极大帮助的历史环境。这里面影响最大的是黎曼19世纪的数学发现，他严格建立了描写任意维弯曲空间的几何方法。1854年在哥廷根大学那篇著名的就职演讲中，黎曼砸碎了平直空间的欧几里得思想锁链，开辟了一条“民主的”几何道路——用统一的数学方法处理各种不同的弯曲空间。正是黎曼的这种思想，为图3.4和图3.6那样的弯曲空间带来了定量的分析方法。爱因斯坦的天才在于他认识到这个数学宝贝仿佛就是为他实现引力新形象而定做的。他大胆宣言，黎曼几何的数学与引力的物理学是天生的姻缘。

然而，在爱因斯坦的绝妙发现约百年后的今天，弦理论为我们提供了一个引力的量子力学图景，不得不在距离尺度小到普朗克长度时修改广义相对论。因为黎曼几何是广义相对论的数学灵魂，所以它也必然需要改变，才可能忠实反映短距离下的弦理论景象。广义相对论断言宇宙的弯曲性质由黎曼几何描述，弦理论则认为只有我们在大尺度下看宇宙才会那样。在普朗克长度那样的小尺度下，一定会出现一种新的几何，来做新的弦理论物理学的伴侣；这门新的几何框架叫量子几何。

与黎曼几何的情形不同，弦理论家找不到什么现成的数学宝贝躺在哪个数学家的书橱里，可以拿来当量子几何。所以，物理学家和数学家们今天正轰轰烈烈研究弦理论，一点点筑成一门新的物理学和数学的分支。尽管完整的故事还有待别人来写，他们的研究已经揭开了许多弦理论所赋予时空的新的几何性质——爱因斯坦见了也会惊愕的性质。

## 黎曼几何

如果你在弹簧垫子上跳，垫子会因你的重量而塌陷、弯曲。陷得最深的是你落脚的地方，而边缘则没多大影响。如果在垫子上画一幅你熟悉的《蒙娜丽莎》，你会清楚地看到这个过程。当弹簧垫子上什么也没有时，蒙娜丽莎与寻常一样；但当你站在垫子上时，画会变形，特别是你脚下的部分，如图10.1。

这个例子触及了黎曼弯曲几何数学框架的根本特征。在高斯（Carl Friedrich Gauss）、罗巴切夫斯基（Nikolai Lobachevsky），波里亚（Janos　Bolyai）等前辈数学家的基础上，黎曼证明了，物体上任何两个位置间的距离可以用来定量表示物体的弯曲程度。粗略地说，不均匀塌陷越大——距离关系偏离平直空间越远——物体的曲率越大。例如，你脚下的垫子陷得最深，在那个区域里两点间的距离关系扭曲也最严重。因此，垫子的这个区域有最大的曲率，这跟你预料的一样。蒙娜丽莎的脸在那儿经历了最严重的扭曲，她那永恒的谜一般微笑的嘴角露出一丝诡异的表情。

爱因斯坦采纳了黎曼的数学发现，为它赋予了精确的物理学意义。我们在第3章讲过，他说明了时空弯曲体现着引力的作用。不过，现在我们要更近地来思考这种解释。从数学上讲，时空曲率——与床垫的弯曲一样——反映了时空点之间的距离关系的扭曲。从物理学看，物体感觉的引力是这种扭曲的直接结果。实际上，如果让物体更小，当我们更深入地认识点的抽象的数学概念的物理意义时，物理

和数学将走得更近。但是，弦理论限制了引力物理学能在多大程度上实现黎曼几何的数学体系，因为它限制了我们能让物体变得多小。当我们走近一根根的弦时，就不能走得更远了。弦理论没有传统的点粒子概念——这是它能为我们带来引力的量子理论的基本因素。这具体说明在根本上依赖于距离概念的黎曼几何结构，在超微观尺度下被弦理论改造了。



图10.1　当你站在“蒙娜丽莎床垫”上时，她的微笑扭曲了。

这些发现对广义相对论的宏观应用没有产生多大影响。例如，在宇宙学中，物理学家依然把星系当作一个个的点，因为它们的大小与整个宇宙比起来是小得可怜的。因此，以这种粗略的方式实现黎曼几何还是很精确的近似，广义相对论在宇宙学背景的成功也证明了这一点。但是，在超微观的领域，弦的延展本性意味着黎曼几何肯定不会是正确的数学形式。正如我们即将看到的，它将被弦理论的量子几何所取代，我们将面临一些崭新的意想不到的特征。

宇宙大舞台

根据宇宙学的大爆炸模型，整个宇宙大约是在150亿年前从一场奇异的大爆炸中轰然出现的。今天，我们看到——最早是哈勃发现的——大爆炸的“碎片”，那亿万个星系，还在向外奔流着。宇宙在膨胀。我们不知道宇宙是一直这样膨胀下去，还是有那么一天会慢慢停下来，然后反过来经历一场宇宙的大收缩。天文学家和天体物理学家正努力从实验来解决这个问题，因为答案引来一个原则上可以观测的量：宇宙的平均物质密度。

假如平均密度超过每立方厘米十万亿亿亿分之一（$10^{-29}$）克的所谓临界密度——相当于宇宙中每立方米中有5个氢原子——那么足够强大的引力将穿透宇宙，把它从膨胀拉回来。假如平均密度比临界值小，引力作用会很弱，挡不住宇宙永不停歇的膨胀。（凭你自己的观察，你大概以为宇宙物质的平均密度远远超过了临界值。但别忘了，物质与金钱一样，会朝着某些地方聚集。拿地球或太阳系，或银河系的物质密度来作为整个宇宙的密度指标，就像拿比尔·盖茨（Bill Gates）的财产来作为全球财富的指标，我们知道多数人的财产与盖茨相比都是微不足道的，平均下来要小得多；同样，星系间存在着大量几乎真空的区域，它们会大大降低宇宙的平均物质密度。）

天文学家通过仔细研究星系在空间的分布，很好掌握了宇宙可见物质的平均量。结果比临界值小许多。但不论从理论还是实验，都有证据强烈表明宇宙还充满着看不见的暗物质。这些物质不参与恒星能源的核聚变，所以不会发光，不能走进天文学家的望远镜。现在还没人能认定暗物质的本性，更谈不上确定它存在的总量。看来，我们还说不清今天膨胀的宇宙会有什么样的命运。

为了讨论方便，让我们假定物质密度真的超过了临界值，在遥远未来的某一天，宇宙将不再膨胀，而开始坍缩。所有星系将慢慢靠近，随着时间的流逝，它们靠近的速度将越来越快，最后以疯狂的速

度撞在一起。你应该看到，整个宇宙在挤压成一块不断收缩的物质。就像我们在第3章讲的那样，它从亿万光年开始收缩，速度在每时每刻增大，万物在不停地汇聚到一起，挤进一个星系大小的空间；它收缩到百万光年，然后到一颗恒星的大小，然后，一颗行星，一个橘子，一颗豆，一粒沙……照广义相对论，它还要继续收缩下去，成一个分子，一个原子，最后在无法抗拒的宇宙大挤压下，它没有了大小。根据传统理论，宇宙从没有大小的原初状态爆炸而来，如果有足够的质量，它又在大收缩中回到相同的终极挤压状态。

但是，我们现在很清楚，当距离尺度在普朗克长度附近或者更小时，量子力学使广义相对论的方程不再有效。我们必须运用弦理论。那么，既然广义相对论允许宇宙的几何形式可以任意小——这相应于黎曼几何的数学允许我们想象任意小的抽象的几何形式——我们自然会问，弦理论是如何改变这种图景的呢？我们很快会看到，弦理论以一种奇异的方式为物理学能达到的距离尺度确立了一个下限，它声称宇宙在任何空间维上都不可能收缩到普朗克长度以下。

这个结果是怎么来的？你可能忍不住要凭自己现在对弦理论的认识，大胆猜想一个答案。当然，你会说不论多少个点堆起来——点粒子就是那样的——体积总还是零。不过，假如粒子真是一根根的弦，它们在完全随机的方向坍缩在一起，就可能填满体积不为零的一团，仿佛一个橡皮筋卷起来的普朗克大小的皮球。如果你这样想，那就走对路了，但可能会忽略一些重要而微妙的特征——弦理论巧妙地利用这些特征发现宇宙应该有一个极限的小尺度；这些特征则具体说明了即将到来的新的弦物理学和它可能给时空几何带来的影响。

为解释这些重要问题，我们先来看一个例子，它忽略了无关紧要的细节，但又不损害新物理学的特征。我们不考虑弦理论的所有10个空间维——甚至我们熟悉的4个展开的时空维也不都考虑——我们还是回到那个花园水管的宇宙。在第8章引进这个二维宇宙是为了说明20世纪20年代卡鲁扎和克莱茵的思想。现在我们用它来作为一个“宇

宙大舞台”。看看弦理论在这样简单的情形会有些什么性质，然后我们根据这样得来的知识去更好地认识弦理论所要求的所有空间维。为达到这个目标，我们想象管子宇宙的横向维度开始是圆鼓鼓的，然后慢慢收缩，圆圈越来越小，管子越来越细，趋向一根直线——这样我们看到一个简化的大挤压过程的缩影。

我们的问题是，这样的宇宙坍缩的几何和物理性质，在弦的宇宙和在点粒子的宇宙间会有什么显著不同的特征吗？

新特征

我们用不着远远地去寻找弦物理的什么基本新特征。一个在二维世界的点粒子可以像图10.2画的那样：在水管伸展的方向运动，在环绕的方向上运动，或者在两个方向之间运动。一根弦的小圈也能这样，不同的是，它会在运动中振动，如图10.3（a）。这点差别我们已经较详细地讨论过了：弦通过振动而产生诸如质量和力荷等特征。虽然这是弦理论的决定性的方面，但我们现在不谈它，因为我们已经懂得了它的物理意义。

我们现在感兴趣的是点粒子运动与弦运动的另一点差别，它直接依赖于运动所在空间的形态。因为弦是展开的物体，所以除了已经讲的那些，还有一种可能的形式：它可以像绳子一样缠绕着管子世界，如图10.3（b）。[1]弦将仍然在管子上滑行和振动，不过是以缠绕的形式运动。实际上，弦可以缠绕管子任何圈（也画在图10.3（b）），一样在滑行中振动。当弦这样卷曲时，我们说它处于缠绕式的运动。显然，缠绕式的运动是弦固有的可能运动形式，点粒子没有对应的状态。我们现在要来认识，这类性质全然不同的运动，对弦本身和它所缠绕的维度的几何性质会产生什么影响。

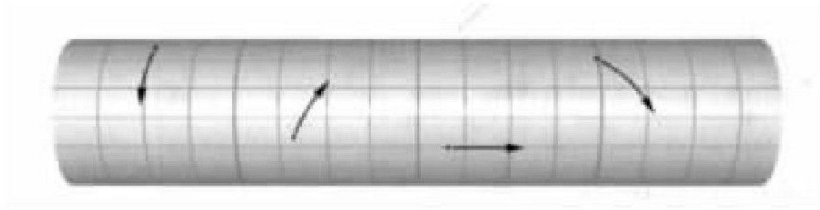


图10.2　在柱面上运动的点粒子。

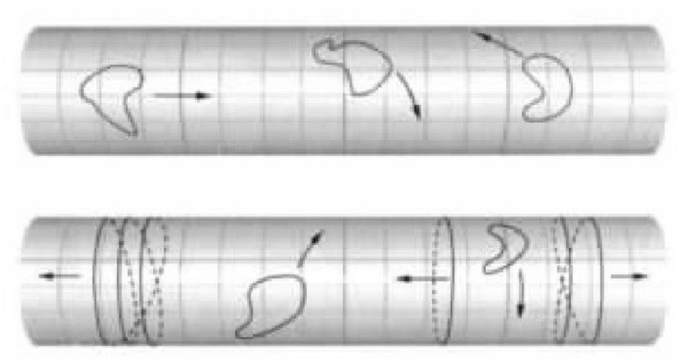



图10.3　弦在柱面上能以两种不同方式运动——“缠绕式的”（b）和“非缠绕式的”（a）。

缠绕的弦

我们前面关于弦的运动讲的都是未缠绕的弦。缠绕空间的圆圈维的弦也几乎都有我们讲过的那些弦的性质。它们的振动也跟未缠绕的伙伴一样，决定着它们的观测性质。两者的基本差别是，缠绕的弦有一个极小质量，取决于卷缩维的大小和缠绕的圈数。弦的振动则决定超过极小质量的那部分质量。

我们很容易明白那极小质量是怎么来的。一根缠绕的弦有极小长度，那是卷缩维的周长和弦缠绕它的圈数所决定的。弦的极小长度决定它的极小质量：弦越长，它的极小质量越大，因为“东西更多”。由于圆周长正比于半径，所以缠绕弦的极小质量正比于缠绕圆周的半径。用爱因斯坦的$E=mc^2$把质量同能量联系起来，我们也可以说束缚

在缠绕弦内的能量正比于被缠绕的圆周的半径。（未缠绕的弦也有极小长度，否则便又回到点粒子的王国了。因为同样的理由，我们说未缠绕的弦也有一个极小但非零的质量。从某种意义说这是对的，但第6章讲的那种量子力学效应（再想想那个“价格游戏”却可能消除这部分质量——零质量的光子、引力子和其他无质量或几乎无质量的粒子就是这样产生出来的。从这点看，缠绕的弦是不一样的。）

缠绕弦的存在如何影响它所缠绕的空间维的几何性质呢？日本物理学家吉川敬治（Keiji Kikkawa）和山崎政实（Masami Yamasaki）在1984年第一次找到一个答案，令人惊讶而困惑。

我们来看管子宇宙大收缩的最后那“惊天动地”的一幕。照广义相对论的方式，卷缩的空间向着普朗克长度收缩，然后继续朝更小的尺度收缩下去；关于这一幕实际发生的事情，弦理论却有着迥然不同的说法。弦理论认为，卷缩维半径小于普朗克长度而且还在减小的管子宇宙所发生的所有物理学过程，与半径大的而且还在增大的管子宇宙所发生的过程，绝对是完全相同的！这就是说，当卷缩的空间向着普朗克尺度和更小的尺度坍缩时，一切的努力都被弦化解了，弦把空间几何扭转回来。弦理论证明，这种演化还可以说成是——或者更准确地解释为——卷缩的空间收缩到普朗克尺度，然后开始扩张。弦理论重写了短距离下的几何定律，原来似乎完全的宇宙坍缩现在好像成了宇宙反弹。卷缩的维可以收缩到普朗克长度，但因为弦的缠绕，再往下收缩实际却成了扩张。我们来看那是为什么。

弦的状态[65]

新的弦的缠绕形式的出现，意味着管子宇宙中弦的能量有两个来源：弦的振动和缠绕。根据卡鲁扎和克莱茵的传统，这两个来源都依赖于管子的几何，也就是说，依赖于卷缩圆圈的半径，不同的是带上了明显的弦的特征，因为点粒子是不可能发生缠绕的。于是，我们的第一件事情是准确地决定弦的振动和缠绕的能量贡献如何依赖于卷缩

维圆周的大小。为此，我们遵照一种被证明是很方便的办法，把弦的振动分解为两个部分：均匀的振动和普通的振动。普通的振动指我们讨论过多次的寻常的振动，如图6.2画的那些振动；均匀的振动说的是一种更简单的运动：弦从一个地方到另一个地方的不改变形状的整体性滑动。所有的弦运动都是滑动与振动的组合，不过在现在的情形，我们很容易把它们区别开来。实际上，普通振动在我们的讨论中不会起多大作用，我们讲完要点以后再考虑它的效应。

我们有两点基本发现。第一点，弦的均匀振动（整体滑动）所激起的能量反比于卷缩维的半径，这是量子力学不确定性原理的直接结果：小半径的空间更严格束缚了弦的活动，从而通过量子力学的幽闭效应增大了弦运动的总能量。所以，当卷缩维的半径减小时，弦运动的能量必然会增大——这明显是反比性的特征。第二点，跟我们以前发现的一样，缠绕运动的能量正比而不是反比于维的半径。记住，这是因为缠绕弦的最小长度——从而也是最小能量——正比于那个半径。这两个事实说明，大的半径意味着大的缠绕能和小的振动能，而小的半径意味着小的缠绕能和大的振动能。

这将我们引向一个重要事实：任何一个卷缩维的圆周半径大的二维世界（或者说较粗的管子世界）都对应着一个半径小的伙伴，前者的弦的缠绕能等于后者的弦的振动能，而前者的弦的振动能等于后者的弦的缠绕能。由于物理学性质关心的是弦结构的总能量——而不在乎能量如何在缠绕和振动间分配——所以这两个几何形态不同的管子世界没有物理学的区别。于是，弦理论得出一个非常令人惊讶的结论：不论管子世界是“粗”还是“细”，它们之间不存在什么区别。

这是宇宙的一个“双赢”策略。假如你是位精明的投资者，你遇到下面的困惑时也会这么做的。假定在华尔街上市的两种股票——一种是做健康器械的，一种是做心脏瓣膜的——牢牢地相互关联着。它们今天的收盘价都是1美元一股。据可靠消息，如果一家股票涨了，另一家就会跌；而且，那位消息灵通人士——他是完全信得过的（尽管

他的做法有点儿违规）——告诉你，明天这两家股票收盘时的价格肯定会互为反比。就是说，如果一家的收盘价是2美元，则另一家该是1/2美元（50美分）；一家是10美元，另一家就是1/10美元（10美分），等等。但是，那人不能告诉你哪家高，哪家低。你该怎么办呢？

你会一下子把所有的钱都投进来，平均分配到两家公司的股票。因为通过几个例子你就能计算出结果，不论第二天股市如何，你都不会赔的。最坏的情形也能保住本钱（两种股票都是1美元1股）；但只要股价有变化——像你的内线说的那样——你总会赚钱的。例如，健康公司在4美元收盘，而心脏瓣膜公司在1/4美元收盘，两者之和是4.25美元，超过了前一天的2美元。而且，从净赚的钱看，你用不着管哪家高哪家低。如果你只关心总的收入，那么两家公司的不同状况并不会对结果发生影响。

弦理论中的能量也处于类似的情形。弦的能量也是两个来源（振动的和缠绕的），两者对总能量的贡献一般是不同的。但我们在下面会看到，不同的几何形态构成的一对——一个产生高缠绕/低振动能，一个产生低缠绕/高振动能——在物理上是没有区别的。另外，在股票的情形，除了总收入以外，两种股票是可以区别的；但两种弦的图景是绝对没有物理学区别的。

实际上，在股票市场也含有类似情形。不过，我们应该考虑另一种投资方式：你没有将钱平均投向两家公司，而是买了1000股健康公司，3000股心脏瓣膜公司。这时候，你的总收入与哪家公司收盘高低有关系。例如，健康收盘10美元，心脏收盘1/10美元时，你原来投入的4000美元现在成了10300美元；如果两家收盘情况相反，则你的股票价值该是30100美元——多多了。

不管怎么说，反比例的股价一定会产生下面的结果。假如你有个朋友，她买股票跟你完全“对着来”——3000股心脏瓣膜公司的，

1000股健康公司的。于是，在健康收盘高（10美元）的情形，她的股值是30100美元，跟你在相反情形的股值一样；同样，当心脏瓣膜收盘高时，她的股值为10300美元，还是跟你在相反情形的股值一样。这就是说，从总的股值看，两个股价的高低更替的影响将完全被两种股票数量的交换所抵消。

记着最后这一点，我们现在回到弦理论，在一个具体例子中考虑可能的弦能量情况。假定管子世界的圆圈半径是普朗克长度的10倍，我们记作R＝10。弦可以缠绕管子任意多圈，如1圈、2圈、3圈，等等。弦缠绕管子的圈数叫缠绕数。缠绕的能量决定于缠绕弦的长度，正比于半径与缠绕数的乘积。另外，任何缠绕的弦都能振动，我们现在讲的整体的均匀振动的能量与半径成反比，也就是半径的倒数1/R（这里是普朗克长度的1/10）的整数倍。我们称这个整数因子为振动数。[2]

你可以看到，这种情形与我们在华尔街遇到的情形很相似。在这里，缠绕数和振动数恰好相应于两家公司股票的份额，而R和1/R则类似于两种股票的收盘价格。现在，我们可以像计算股值那样，通过缠绕数、振动数和半径来计算弦的总能量。表10.1列举了部分弦状态的总能量。表中还列举了在管子半径R＝10情况下我们选择的缠绕数和振动数。

缠绕数和振动数可以是任何整数，所以完整的表是无限长的。不过，就我们的讨论来说，这几行有足够代表性。从表中可以看到，我们选择的是高缠绕能和低振动能的状态：缠绕能的因子为10，而振动能的因子为1/10。

现在想象管子收缩，半径从10缩到9.2、7.1……直到1.1、0.7，最后收缩到0.1（1/10），停下来。我们现在讨论这种情形。对这个几何特征不同的管子宇宙，我们可以得到类似的一个弦能量表：现在缠绕能的因子是1/10，而振动能的因子是它的倒数10。结果是表

10.2。

| 振动数 | 缠绕数 | 总能量 |
| --- | --- | --- |
| 1 | 1 | 1/10＋10＝10.1 |
| 1 | 2 | 1/10＋20＝20.1 |
| 1 | 3 | 1/10＋30＝30.1 |
| 1 | 4 | 1/10＋40＝40.1 |
| 2 | 1 | 2/10＋10＝10.2 |
| 2 | 2 | 2/10＋20＝20.2 |
| 2 | 3 | 2/10＋30＝30.2 |
| 2 | 4 | 2/10＋40＝40.2 |
| 3 | 1 | 3/10＋10＝10.3 |
| 3 | 2 | 3/10＋20＝20.3 |
| 3 | 3 | 3/10＋30＝30.3 |
| 3 | 4 | 3/10＋40＝40.3 |
| 4 | 1 | 4/10＋10＝10.4 |
| 4 | 2 | 4/10＋20＝20.4 |
| 4 | 3 | 4/10＋30＝30.4 |
| 4 | 4 | 4/10＋40＝40.4 |

表10.1　在图10.3所示宇宙中运动的弦振动和缠绕的例子，缠绕维的半径为R＝10。振动能的因子为1/10，缠绕能的因子为10，从而得出所列的总能量。能量单位为普朗克能量。例如，表中最后一列10.1的意思是10.1倍普朗克能量。

乍看起来，两张表是不同的。但仔细看看，除了数字的次序不同外，两表的“总能量”是完全相同的。为在表10.1中找到与表10.2的某个能量对应的值，只需要交换缠绕数和振动数。就是说，当卷缩维的半径发生改变时（如从10到1/10），振动与缠绕所扮演的角色也相互替换了。于是，只要我们考虑弦的总能量，卷缩维的大小就不会产生什么影响。像那两种股票价格的变化完全被股票份额的交换所补偿一样，把半径从10调换为1/10的结果，也将通过交换振动数和缠绕数而消化。而且，这种结论对任何半径和它的倒数都是成立的，我们选择R＝10与R＝1/10不过是为了简单方便。[3]

| 振动数 | 缠绕数 | 总能量 |
|---|---|---|
| 1 | 1 | 10＋1/10＝10.1 |
| 1 | 2 | 10＋2/10＝10.2 |
| 1 | 3 | 10＋3/10＝10.3 |
| 1 | 4 | 10＋4/10＝10.4 |
| 2 | 1 | 20＋1/10＝20.1 |
| 2 | 2 | 20＋2/10＝20.2 |
| 2 | 3 | 20＋3/10＝20.3 |
| 2 | 4 | 20＋4/10＝20.4 |
| 3 | 1 | 30＋1/10＝30.1 |
| 3 | 2 | 30＋2/10＝30.2 |
| 3 | 3 | 30＋3/10＝30.3 |
| 3 | 4 | 30＋4/10＝30.4 |
| 4 | 1 | 40＋1/10＝40.1 |
| 4 | 2 | 40＋2/10＝40.2 |
| 4 | 3 | 40＋3/10＝40.3 |
| 4 | 4 | 40＋4/10＝40.4 |

表10.2　同表10.1，但半径为R＝1/10。

表10.1和表10.2是不完整的，原因有两个。第一个我们讲了，弦的振动数和缠绕数可以有无限多的可能，而我们只列举了几个。这当然不会有什么问题——我们只要有耐性，想把表列多长都行。我们会发现，表中的关系总是成立的。第二个原因是，除缠绕能外，我们只考虑了来自弦的均匀振动的能量。现在，我们要把普通振动也考虑进来，它们为总能量带来另一份贡献，而且还决定着弦携带的力荷。但更重要的是，这些贡献与半径大小无关。这样，即使我们在表10.1和表10.2里考虑了这些更具体的特性，两个表还是相互对应的，因为普通振动的贡献在任何情况下都是相同的。于是，我们可以说，半径为R的管子世界里粒子的质量和力荷与半径为1/R的情形是完全一样的。因为质量与力荷决定着基本的物理现象，所以在物理上我们不能区别这两种不同几何的宇宙。一个宇宙做的实验在另一个宇宙中有一个对应的实验，它们将导出相同的结果。

争论

乔治和格雷茜走进二维管子世界，成了二维生命，做了那里的物理学教授。两人各建起一个与对方竞争的实验室，都宣布自己确定了卷缩维的半径。两人的实验精度一贯令人佩服，但奇怪的是这回他们的结果却是矛盾的。乔治说半径R＝10倍普朗克长度，而格蕾茜宣称R＝1/10倍普朗克长度。

“格蕾茜，”乔治说，“据我的弦理论计算，我知道，假如圆圈维的半径是10，我就能预期看到表10.1所列的那些能量的弦。我已经用新的普朗克能量加速器做了好多实验，已经证实了这个预言。所以，我相信，我敢说那圆的半径是R＝10。”格蕾茜替自己说了差不多同样的话，不过她的结论是她发现了表10.2所列的能量，从而证明半径R＝1/10。

格蕾茜灵机一动，让乔治看到两个表虽然次序不同，内容却是完全一样的。可乔治总要迟钝一些，他问，“怎么会这样呢？根据量子力学和缠绕弦的基本特征，我知道不同的半径会产生不同的弦能量和力荷，如果承认这一点，那我们的半径应该是相同的。”

格蕾茜根据她对弦物理学的新发现告诉乔治：“你说的差不多是对的，可不完全。一般情况下，不同的半径会产生不同的能量；但在特殊情形，例如两个半径互为倒数——10与1/10——则允许的能量和力荷实际上是完全一样的。你看，你说的缠绕，我说是振动，而你说的振动，我说是缠绕。大自然可不管我们怎么说，物理学决定于基本的物质构成——粒子质量（能量）和它所带的力荷。不论半径是R还是1/R，弦理论中基本物质构成的这些性质是完全一样的。”

乔治费好大气力才明白过来，他回答说：“我想我明白了。虽然你我给出的弦的具体描述有所不同——要么缠绕卷缩维的方式不同，要么振动行为不同——但它们表现的物理学特征却是完全相同的。因

为宇宙的物理学性质依赖于这些基本物质组成的性质，所以在半径互为反比的两个宇宙间没有什么不同，也没有办法区分它们。”说得完全正确。

三个问题

现在你可能会问：“你看，假如我是管子世界里的一个小生命，我可以很简单地拿皮尺去测量管子的周长，从而毫无疑问地确定它的半径——没有假设，也没有但是。那么，不同半径而又不可分辨的两个世界有什么意思呢？另外，弦理论不是排除了普朗克长度以下的尺度了吗，为什么我们还在谈多少分之一普朗克长度的半径的维度呢？最后，虽然我们在讲二维的管子世界，但谁会把它当真呢？——当我们把所有的维都考虑进来时，它还能有什么意义吗？”

我们先来看最后这个问题，答案会把我们引向前两个问题。

虽然我们在二维管子世界里进行讨论，仅限于1个展开维和1个卷缩维，但这样做只是为了简单。如果我们有3个展开维和6个卷缩维——后者是所有卡-丘空间里最简单的形态——那些结论也是完全一样的。每个卷缩维有一个半径，它与半径为倒数的维将生成在物理学上完全相同的宇宙。

我们甚至还可以把这个结论推得更远。在我们的宇宙中，可以看到三个展开的空间维，据天文学家的观测，它们看起来都延伸到大约150亿光年（1光年大约是9万亿千米，所以这延伸的距离大概是1.4亿亿亿千米）。我们在第8章讲过，没人知道那距离以外在发生什么。我们不知道它们是继续无限延伸下去，还是把自身卷缩成超出我们望远镜“感觉能力”的一个巨大的圆。假如它们是卷曲的，那么在太空远行的宇航员不断朝着同一个固定的方向走下去，就能最终绕宇宙一圈——像麦哲伦（Magellan）环游地球那样——回到原来出发的地方。

看来，我们熟悉的展开维也可能是些圆圈，从而也像弦理论说的那样，R与1/R的世界是不可区别的。具体说，这些圆的半径应该是刚才讲的150亿光年，是普朗克长度的10万亿亿亿亿亿亿亿（$10^{61}$倍），而且还在随宇宙膨胀而增大。如果弦理论是对的，这个宇宙与一个展开维的半径只有$1/R = 1/10^{61} = 10^{-61}$普朗克长度的宇宙在物理学上是一样的！这是在弦理论下我们熟悉的宇宙空间的另一幅图景。实际上，在那个“倒数世界”，小圆圈还将随时间变得更小。因为R增大，1/R自然会缩小。现在我们似乎真的走到尽头了。这能是真的吗？我们六英尺的身躯怎么可能“活”在这样难以置信的微观世界里？那么“一丁点儿”宇宙怎么能在物理上与我们看到的茫茫太空相同呢？而且，我们现在也自然走近上面提的第二个问题：弦理论似乎剥夺了我们探索普朗克尺度以下的距离的能力。但是，假如圆半径R大于普朗克长度，它的倒数1/R自然只有普朗克长度的若干分之一。那么结果呢？答案将关联我们的第一个问题，而且揭示了空间和距离的重要而奇妙的一面。

两个距离

距离是我们认识世界的一个十分基本的概念，似乎很简单，人们常常忽略它还有玄之又玄的地方。狭义和广义相对论曾给我们关于空间和时间的概念带来过惊人的影响，弦理论也生出一些新奇的特征，这些经历使我们今天在距离的概念上也更小心翼翼了。物理学中最有意义的定义是那些可操作的——就是说，定义为所定义的东西提供了至少是原则上的测量方法。毕竟，不管概念如何抽象，有了可操作的定义，我们就能在实验中揭示它的意义，测量它的大小。

我们如何才能得到一个可操作的距离的定义呢？在弦理论的背景下回答这个问题会令人大吃一惊的。1988年，布朗大学的布兰登伯格（Robert Brandenberger）和哈佛大学的瓦法两位物理学家指出，假如某个空间维是圆，那么在弦理论中存在着两个不同然而相关的可操作定义。每个定义都有一套不同的测量距离的实验程序，而测量的基

础大致说来却是一个很简单的原理：如果探针以已知固定的速度运动，我们可以根据它经过某个距离的时间来确定那段距离的长度。两个定义的差别在于实验过程所选择的探针不同。第一个定义用的是未缠绕在圆圈维的弦，而第二个定义用的是缠绕的弦。我们看到，弦理论中存在两种不同的可操作的距离定义，原因正在于所用的基本探针具有延展的本性。在点粒子理论中没有缠绕的概念，所以只有一种距离定义。

两种操作过程会有怎样不同的结果呢？布兰登伯格和瓦法的发现既令人惊奇，也难以捉摸。借助于不确定性原理，我们大概能明白那答案的意思。未缠绕的弦可以自由沿着圆周滑动，长度正比于半径R。根据不确定性原理，弦的能量正比于1/R（回想一下我们在第6章讲过的探针的能量与它对距离敏感性的关系）。另一方面，我们知道缠绕的弦有着正比于R的极小能量，于是不确定性原理告诉我们，它对距离的敏感程度正比于R的倒数，1/R。将这个思想用数学公式表达出来，我们就能看到，如果拿它们来测量空间的卷缩维的半径，那么未缠绕的弦将测得R，而缠绕的弦将测得1/R。这里，我们的测量还是像从前一样，以普朗克长度为单位。两个实验都可以说自己的结果是圆周的半径——弦理论教导我们的是，以不同探针来测量距离可以得到不同的结果。实际上，这个性质可以推广到所有长度和距离的测量，而不仅限于确定卷缩维的大小。缠绕与未缠绕的弦探针所获得的结果将互成反比。[4]

如果宇宙真像弦理论描绘的那样，我们为什么没在寻常的生活和科学活动中遇到过那两种可能的距离概念呢？我们讲距离的时候，似乎总是从经验来讲的，只有一种距离，没有任何线索暗示还藏着另一种距离的概念。我们为什么会错过那个可能呢？原来，尽管在我们的讨论里R与1/R是高度对称的，但当R（从而1/R也）远远偏离1（当然还是指1个普朗克长度）时，两个可操作的定义中有一个是极难实现的（虽然还有一个是极易实现的）。大概说来，我们总是操作那个容

易的，完全不知道还有另一种可能。

两种方法难易悬殊的原因在于所用探针的质量大不相同——要么缠绕的能量高，要么振动的能量高。假如半径R（从而1/R也）远离普朗克长度（即R＝1），这时候，所谓“高”能相应于重得惊人的探针——例如比质子重百亿亿倍，而所谓“低”能，差不多就是比零质量重一点儿的探针。在这样的背景下，两种方法便有着天壤的难易差别。因为，光是产生那样的重弦形态也远远超越了我们今天的技术能力。因此，在实践中，只有那个涉及较轻的弦形态的方法才有技术上的可能，那也是我们在讨论距离问题时一贯用的方法。这种方法培养了我们的直觉，从而也符合我们的直觉。

把实际抛到一边，在弦理论主宰的宇宙中，我们可以自由选择一种方法来测量距离。天文学家测量“宇宙的大小”，是通过检验穿过太空碰巧进入他们望远镜的光子；显然，光子在这儿可真是光光的没有质量的弦。结果，光子测得的距离是$10^{61}$倍普朗克长度，前面已经说过了。假如我们习惯的那3个空间维也是卷缩的，假如弦理论是正确的，那么从原则上讲，用迥然不同的（当然现在还没有的）仪器的天文学家，应该能测量重弦缠绕的空间有多大，他们将发现那距离是光子测得距离的倒数。在这个意义上，我们可以认为宇宙既可能像我们寻常感觉的那么大，也可能小得可怜。根据轻弦模式，宇宙是巨大而膨胀的；而据重弦模式，宇宙是渺小而卷缩的。这里没有矛盾，而是存在着两种不同然而却同样合理的距离定义。由于技术的限制，我们很熟悉第一种定义，而不管怎么说，两个概念都是一样有效的。

现在我们来回答前面的问题，大人如何能在小宇宙中生存？当我们测量一个人的身高，说他高1.8米时，我们一定在用轻弦模式。为比较他们与宇宙的大小，我们必须用同样的过程来测量宇宙，上面说过，那是150亿光年，比1.8米大多了。这样的人类如何能活在重弦模式所测量的“小”宇宙中呢？这是一个没有意义的问题——是在拿苹果同橘子比。现在我们有了两个距离概念——轻弦探针的和重弦探针的

——我们也该在相同的模式下比较测量结果。

最小尺度

慢慢往前走，我们就要到头了。如果我们坚持用“容易的办法”来测量——也就是用最轻的弦模式来测量——结果将总是大于普朗克长度。为看清这一点，我们考虑假想的三维空间的大收缩，并假定我们熟悉的那三维是圆的。为讨论方便，假定在思想实验的开始，未缠绕的弦模型是轻的，我们用它来测量宇宙，发现它有一个巨大的半径，正在随时间而收缩。当它收缩时，未缠绕的弦变得越来越重，而缠绕的弦越来越轻。当半径一路收缩到普朗克长度——即R＝1时——缠绕的弦与振动的弦正好有相同的质量。这时，两种测量距离的方法都同样难以实现；而且，它们将得出相同的结果，因为1也是它自己的倒数。

半径继续往下收缩，缠绕的弦将变得比未缠绕的更轻，这样，它们自然成为我们用以测量距离的“更容易的方法”。根据这种测量，结果是较重的未缠绕弦的结果的倒数，即半径大于1个普朗克长度，并且还在增大。这不过反映了，当未缠绕弦测量的R收缩到1，并继续收缩时，缠绕弦所测量的1/R将增大到1并且继续增大。于是，当我们决意总以轻弦模式这种“更容易”的方法来测量距离时，我们遇到的最小半径就是普朗克长度。

因为轻弦模式测量的宇宙半径总是大于普朗克长度的，一个特别的结果就是，我们避免了一个会趋向于零的大收缩。根据最轻弦模式的测量，宇宙半径不会朝比普朗克长度更小的方向收缩，当它收缩到普朗克长度时，它会反过来开始增大。反弹的一幕替代了无限的大挤压。

用轻弦模式测量距离，符合我们关于长度的传统概念——那是早在弦理论发现以前就形成的了。如我们在第5章看到的，正是因为这

个距离概念，我们才在普朗克尺度以下的距离遇到了不可克服的剧烈的量子波澜。我们又一次看到，弦理论凭它的两个互补的距离概念避免了那可怕的超短距离。在广义相对论的物理学框架和相应的黎曼几何的数学框架下，距离的概念只有一个，它可以是任意小的数值。在弦理论的物理学框架和相应的新生的量子几何的领域里，距离的定义有两个。小心翼翼地运用这两个定义，我们发现有一个概念在大尺度下，与我们的直觉和广义相对论都是相容的，但在小尺度下却迥然不同。具体说来，小于普朗克尺度的距离是不可能达到的。

上面讲的有点儿玄，我们把关键的一点再强调一遍。假如我们硬是不在乎什么“难”与“易”的距离测量方法，而要坚持用未缠绕的弦来测量，那么当R收缩到普朗克长度以下时，我们似乎真能走近比普朗克尺度更小的距离。但上面的讨论告诉我们，那所谓“更小的距离”需要小心来理解，因为它可以有两种不同的意思，而只有一种符合我们的传统观念。在这里，当R收缩到普朗克长度以下时，如果我们还坚持用未缠绕的弦（这时它们已经变得比缠绕的弦更重了），那我们实际上是在用“难”的方法来测量距离，从而那“距离”的意思不满足我们标准的用法。然而，这里讨论的绝不仅仅是语义学的问题、传统习惯的问题或者测量的可行性问题。即使我们愿意用非标准的距离概念来描写一个比普朗克长度更小的宇宙，我们遇到的物理学——如前几节讨论的——并没有什么不同，还是那个大半径宇宙（传统距离的表义下）的物理学（举例来说，就像表10.1与表10.2之间对应的物理学）。真正有意义的正是物理，而不是语言。

布兰登伯格、瓦法和其他一些物理学家根据这些思想重新写下了宇宙学定律，在那里，大爆炸和可能的大收缩都不再牵扯一个零尺度的宇宙，而是每个维都是普朗克长度的宇宙。这当然是一个诱人的图景，原来那个起源于并可能坍缩成一个无限致密的点的宇宙所具有的那些数学的、物理的和逻辑的难题都烟消云散了。尽管很难想象整个宇宙卷缩在一个普朗克尺度的小球里，但比起想象它挤压成一个没有

大小的点，还是好得多了。我们将在第14章讨论，弦宇宙学还是一个年轻的领域，不过希望很大，很可能为我们带来这样一个比标准大爆炸模型更容易理解的模型。

## 结论普遍吗

如果空间维不是圆，结果会怎样呢？那些关于弦理论的最小空间距离的惊人结论还能成立吗？谁也说不准。圆形维度最基本的特征是允许弦的缠绕。只要空间维——不论什么形状——允许弦的缠绕，我们讲的大多数结论应该还是成立的。但是，假如有两维是球形的呢？这种情况下弦不能“牢牢”绕在球面上，因为它总会“滑落下来”，像一根橡皮筋从篮球上滑下来。另外，弦理论限定了这些维的收缩尺度吗？

大量研究表明，答案有赖于一个完全的空间维是卷缩的（如我们这一章讲的），还是在坍缩的孤立的“一小块”空间（我们将在第11章和第13章讨论）。弦理论家普遍相信，不论形状如何，只要我们是在让一个完整的空间维发生收缩，它就像圆的情形一样，有一个极限的尺度。确立这一观念是未来研究的一个重要目标，因为它将直接影响弦理论的诸多方面，包括它的宇宙学意义。

## 镜像对称

爱因斯坦通过广义相对论在引力的物理学与时空的几何学之间建立了联系。乍看起来，弦理论巩固并拓宽了物理与几何的联系，因为振动弦的性质——它们的质量和所携带的力荷——基本上取决于空间卷缩部分的性质。不过，我们刚开始看到了，量子几何这一弦理论的几何物理学还有些奇奇怪怪的东西。在广义相对论和“传统”几何学中，半径为R的圆与半径为1/R的圆是绝对不同的；然而在弦理论中，它们在物理上却是不可区别的。这使得我们敢雄心勃勃地往前走得更远，我们想，也许还有差别更大的空间几何形式——不仅大小不同，

形态也不同——但在弦理论中却找不出它们有什么物理的差别。

1988年，斯坦福大学直线加速器中心的狄克松（Lance Dixon）有一个关于这方面的重大发现，欧洲核子中心（CERN）的勒克（Wolfgang Lerche）、哈佛的瓦法和当时在麻省理工学院（MIT）的瓦纳（Nicholas Warner）也发现了同样的东西。这些物理学家在基于对称性考虑的美学原则下提出一个大胆猜想：为弦理论的卷缩维选择的两种不同卡-丘空间，也许能生成相同的物理。

为说明这种奇异的可能性如何能够发生，我们回想一下，卡-丘空间的孔洞数决定着弦能产生多少族可能的激发态。这些孔洞像我们见过的单孔或多孔的环，如图9.1。我们在纸上画的二维图有一大缺点，不能表现一个六维的卡-丘空间可以具有不同维的孔洞。虽然我们画不出这些孔，但可以用大家理解的数学来描述它们。关键的一点是，源自弦振动的粒子族的数目只依赖于孔的总数，而与某个维的孔数无关（因此，我们在第9章的讨论里并不在意孔的类型有什么不同）。接下来我们想象两个卡-丘空间，它们在不同的维有不同数目的孔，但孔的总数却是相同的。由于不同维的孔数不同，所以这两个卡-丘空间有不同的形态。但因为孔的总数相同，所以它们生成的宇宙有相同数目的粒子族。当然，这不过是一个物理性质。如果要一切物理性质都相同，那要求就严格得多。不过，这一点性质至少能说明狄克松-勒克-瓦法-瓦纳猜想很可能是对的。

1987年秋，我来到哈佛大学物理系做博士后，我的办公室刚好在瓦法的走廊下面。我的学位论文研究的就是弦理论中卷缩卡-丘空间的物理和数学性质，所以瓦法常向我通报他在这方面的工作。1988年秋的一天，他经过我办公室时停下来告诉我，他和勒克、瓦纳有了那个猜想。我很感兴趣，但也有些怀疑。兴趣来自这样的认识：如果猜想是对的，它将在弦理论的研究中开辟一条新路；而怀疑来自我的担心：猜想是一回事，证实那些理论性质却是另一回事。

在接下来的几个月里，我总在考虑他的猜想。坦白地说，我一半认为它是错的。然而，奇怪的是，我与普里泽（Ronen Ples-ser）做过的一个看似不相干的项目令我很快完全改变了看法。普里泽那时是哈佛的研究生，现在在魏茨曼研究所和杜克（Duke）大学，我们曾满怀热情地想发展一种方法，从一个初始的卡-丘空间出发，用数学操作生成一种尚未知晓的卡-丘形态。我们特别感兴趣的是所谓的轨形变换（orbifolding）技术，先前是由狄克松、哈维（Jeffrey Harvey，在芝加哥大学）、瓦法和惠藤在20世纪80年代中期发展起来的。粗略地讲，就是将原来的卡-丘空间里不同的点黏在一起，按一定的数学法则生成一个新的卡-丘空间。图10.4示意了这样一个过程。这幅图背后的数学是很可怕的，因为这一点，弦理论家只是对最简单的空间形态——如图9.1的高维多孔面包圈——考察了这种技术的应用情况。不过，普里泽和我发现，革普纳（Doron Gepner，那时在普林斯顿大学）的一些美妙发现也许能提供一个有力的理论框架，把轨形变换技术推行到如图8.9那样复杂的卡-丘空间。

经过几个月的紧张探寻，我们得到一个令人惊讶的结果。如果以恰当方式把某些特殊的点黏接在一起，生成的卡-丘空间将以一种奇异的方式表现出与原来空间的区别：新空间的奇数维的孔数等于老空间的偶数维的孔数，反过来也对。特别的是，这意味着孔的总数——从而粒子族的数目——是相同的，尽管两个奇偶相对的空间形态和基本的几何结构当然是完全不同的。[5]

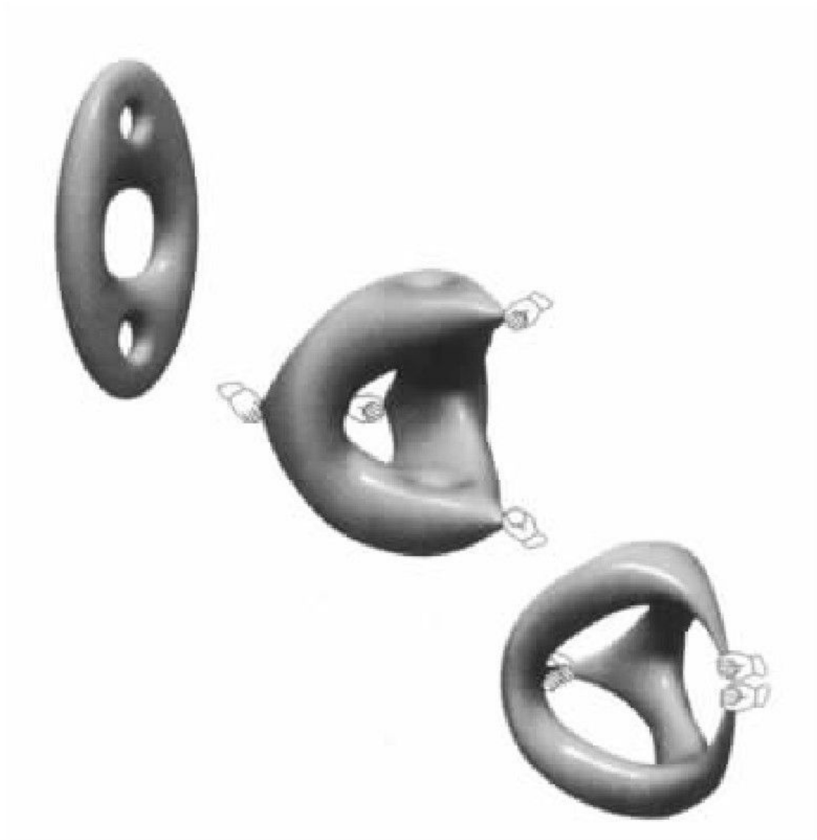


图10.4　所谓环形变换技术是这样一个过程：通过将初始卡-丘空间的不同点黏接在一起而生成一个新的卡-丘空间。

结果显然与狄克松等人的猜想相关，这令我们很兴奋。接下来，普里泽和我又去研究一个关键问题：那两个卡-丘空间除了粒子族的数目相同而外，别的物理性质也相同吗？经过两个多月仔细而艰难的数学分析，其间还得到我的学位论文导师、牛津大学的罗斯（Graham Ross）和老朋友瓦法的启发和鼓励，普里泽和我最后得到了答案：差不多可以肯定是那样的。因为一个与交换奇偶性有关的数学理由，我们以镜像流形来称这些在物理上等价而几何形态不同的卡-丘空间。[6]每一对镜像卡-丘空间当然并不是我们平常讲的字面意义的镜像。尽管它们有不同的几何性质，在用于弦理论的额外空间时，却能生成同一

个物理的宇宙。

发现这个结果后的几个星期，我们是在焦虑中度过的。普里泽和我都明白，我们正在弦理论的一个浪头上，我们证明，爱因斯坦建立的几何与物理学的紧密联系在弦理论中焕然一新了：在广义相对论中意味着不同物理性质的不同几何形态，在弦理论中却可能生成相同的物理。但是，假如我们错了呢？假如那些物理性质以我们忽略了的某种微妙方式产生变化呢？我们把结果告诉了丘成桐，他礼貌然而严厉地指出，我们一定在哪儿错了；他说，从数学观点看，我们的结果太离奇了，不会是真的。他的意见使我们很犹豫。如果一个小结论或不会太引人注意的结论，犯点儿错误也许还算不得什么；而我们的结果是在一个新方向上迈出的意想不到的一步，当然会引起强烈反响。如果它错了，所有的人都会知道。

最后，我们把文章反复检查了，越来越有信心，就拿出去发表。几天以后，我正坐在哈佛的办公室时，电话响了。那是得克萨斯大学的坎德拉斯打来的。他开口就问我是不是坐好了。当然。接着他告诉我，他和两个学生林克（Monika Lynker）和施姆里克（Rolf Schimmrigk）发现了一样东西，会让我从椅子上蹦起来。他们仔细考察了计算机生成的大量卡-丘空间例子，发现这些空间几乎都是成对出现的，两个空间的差别仅在于奇数维和偶数维的洞的数目相互交换了。我告诉他，我还坐得好好的——普里泽和我已发现了相同的结果。坎德拉斯和我们的结果原来是互为补充的：我们走得远一点，证明了镜像空间生成的物理学是一样的；而坎德拉斯和他的学生证明大量的卡-丘空间都以镜像对的形式出现。通过这两篇文章，我们发现了弦理论的镜像对称。[7]

镜像对称的物理学和数学

爱因斯坦在空间的几何与物理的现象间建立的刚性而惟一的联系，在弦理论中获得了解放，这是一个惊人的“范式的转移”。但这些

发展所带来的远不只是哲学态度的改变。镜像对称还特别为认识弦理论的物理和卡-丘空间的数学提供了强大的工具。

在所谓代数几何领域从事研究的数学家，在弦理论发现很久以前就一直在为纯数学的理由研究卡-丘空间。他们发现了这些空间的许多具体性质，没有一个显得有未来的物理意义。不过，卡-丘空间的某些性质已经证明是很困难的——基本上不可能完全揭示出来。但弦理论的镜像对称的发现极大改变了这种局面。大致说来，镜像对称说的是，原来认为毫不相干的特殊的卡-丘空间对现在被弦理论紧紧联系在一起了。联结它们的是一个共同的物理宇宙，任选一个空间作为卷缩维的空间形式，都将生成这样的宇宙。这种意外的内在联系提供了一个新的有力的物理学和数学工具。

举例来说，假如你在忙着计算与卷缩维的某个卡-丘形式相关联的物理性质——如粒子的质量和力荷。你并不特别关心计算结果与实验的联系，因为我们已经看到，现在做那些实验还有大量理论和技术的障碍。实际上，你是在靠思想实验做计算，关心的是假如选择了某个卡-丘空间，宇宙应该像什么样子。开始计算的时候，一切都还顺利；但接着，你的计算遇到了难以逾越的障碍。没人能帮你，世界上最好的数学家也不知道该怎么往下算。你迷失了方向。但是你后来发现这个卡-丘空间有一个镜像伙伴。因为这两个空间生成的弦物理是完全相同的，你意识到自己可以自由地随便拿一个来做计算。于是，你用原来那个卡-丘空间的镜像伙伴重新做刚才那些艰难计算，你相信计算的结果——物理——应该是一样的。起初你可能认为重新做的计算也会像原来那么难，但你却惊喜地发现，两个计算虽然结果会是一样的，但具体形式却大不相同。原来的某些可怕的计算，在镜像的卡-丘空间里变得非常简单了。为什么会这样呢？这不是两三句话就能说明白的，不过，至少对某些计算来说，几乎肯定是这样的，而且计算的难度可以大大降低。它的意义自然是清楚的：你从迷失的方向里走出来了。

这多少有点儿像下面的例子。假设有人陪你数一堆橘子，橘子随便堆放在一只大果箱里，那箱子3米长，3米宽，3米高。起初，你一个个地数，但很快发现这太累人了。幸运的是，这时来了一个朋友，他是看到橘子送来的。他告诉你，橘子原来整整齐齐堆放在小箱子里（他正好拿着一只那样的箱子），小箱子堆在一起，长、宽、高都是20个。你很快算出，送来了8000小箱橘子。现在你需要知道的只是数清一只小箱子里能堆放多少橘子。这是很容易的。你从朋友那儿借来小箱子，用橘子把它填满，这样，原来那艰巨的使命不费吹灰之力就完成了。总之，发现一种聪明的计算方法，做起来就容易得多。

弦理论中的许多计算都是这种情形。从一个卡-丘空间看，计算可能牵涉大量艰苦的数学步骤；然而，如果转移到它的镜像空间，计算可以更有效地重新组织，从而能够相对容易地实现。这一点是普里泽和我发现的，后来，坎德拉斯和他的合作者得克萨斯大学的奥莎（Xenia de la Ossa）、帕克斯（Linda Parkes）和马里兰大学的格林（Paul Green），令人惊奇地将它投入了实践。他们证明，几乎所有困难的计算都能在镜像空间里实现，只需要一台电脑和几页的代算计算。

这对数学家来说更是特别激动人心的发现，因为其中的某些计算也曾令他们困惑过多年。用物理学家的话说，弦理论把它们都解决了。

现在你该明白，在数学家和物理学家之间存在着许多有益的而且通常是友好的竞争。事实上，两个挪威数学家——埃林斯鲁德（Geir Ellingsrud）和斯特罗姆（Arild StrOmme）——就曾计算过坎德拉斯和他的伙伴们用镜像对称成功解决了的一个问题。大体说来，那相当于计算在某个特别的卡-丘空间里能“堆放”多少个球，有点儿像我们在大果箱里数橘子的问题。1991年，在伯克莱举行的一次物理学家和数学家会议上，坎德拉斯宣布他的小组用弦理论和镜像对称得出的结果是317206375。埃林斯鲁德和斯特罗姆也宣布了他们艰难的数学计算

结果：2682549425。几天里，数学家和物理学家一直在争论：谁是对的？这个问题成了弦理论定量可靠性的真正考验。许多人甚至说——多少带点儿玩笑——这是弦理论能否与实验对比的最好检验。另外，坎德拉斯的结果远不仅是埃林斯鲁德和斯特罗姆也计算了的数值结果，他们还宣布说回答了许多别的极端困难的问题——实际上，那些难题连数学家也从未想过。但弦理论可信吗？数学家和物理学家们在会上进行了广泛的交流，可分歧最终还是没能解决。

大约一个月过后，一封电子邮件在参加过伯克莱会议的人中间传开了，信的主题是物理学赢了！埃林斯鲁德和斯特罗姆在他们的计算机代码中发现了一个错误，改正以后他们也证实了坎德拉斯的结果。从那以后，许多数学家都来检验弦理论镜像对称的定量可靠性：所有的检验都胜利通过了。在物理学家发现镜像对称近10年后，最近，数学家在揭示其内在的数学基础方面取得了重大进展。根据数学家康泽维奇（Maxim Kontsevich）、曼宁（Yuri Manin）、田刚（Gang Tian）、李军（Jun Li）和吉温托尔（Ale-xander Givental）等人的重要成果，丘成桐和他的合作者刘克峰等终于从数学上严格证明了用来计算卡-丘空间能放多少个球的公式，从而解决了困扰数学家几百年的一大难题。

除了这场独特的胜利，这些发现真正让我们看到物理学开始在数学舞台上崭露头角了。过去许多时候，物理学家曾在数学的仓库里“发掘”出一些工具来构造和分析物理世界的模型。现在，通过弦理论的发现，物理学家开始偿还他们的债务，为数学提供新的方法去解决他们的未解问题。弦理论不仅树起一个统一的物理学框架，还可能实现一个同样深刻的数学大联合。

注释

1.为了讨论的完整，还应该说明，虽然到现在为止我们在书中讲的许多东西都同样适用于开弦（两端自由的弦）或闭弦圈（这正是我

们所关心的），但在这里讨论的问题上，两种弦将表现出不同的性质。毕竟开弦是不会缠绕在某个卷缩维的。不过，圣巴巴拉加利福尼亚大学的Joe Polchinski和他的两个学生戴建辉（Jian-Hui Dai）和Robert Leigh在1989年说明了开弦如何能很好地符合我们在这一章得到的结论。他们的成果最终将在第二次超弦革命中发挥重要作用。

2.如果你想知道为什么均匀振动的允许能量是1/R的整数倍，请回想一下第4章的量子力学讨论——特别是关于那个仓库的讨论。我们从那里知道，量子力学的能量像钞票一样，是离散的能量“元”组成的：是不同能量“元”的整数倍。在管子世界均匀振动的弦的情形，能量元正好是1/R，我们在正文里用不确定性原理解释过了。这样，均匀振动的能量就是1/R的整数倍。

3.从数学上讲，在卷缩维半径为R或1/R的宇宙中，弦能量的形式为v/R＋wR，这里v为振动数，w为缠绕数。同时交换v与w和R与1/R——即交换振动数与缠绕数，同时半径换为倒数，这个方程的形式是不变的。这就是两个宇宙的弦能量相同的原因。我们在讨论中用的是普朗克单位，也可以换成更传统的单位，即用一个所谓弦标度 $\sqrt{\alpha'}$ 来改写能量公式——弦标度的值大约是普朗克长度，$10^{-33}$厘米。这样，弦能量可以表达为v/R＋wR/α′，在交换v与w和R与α′/R时，它是不变的。这里，R和α′/R用的是传统的距离单位。

4.你大概很奇怪，在半径为R的卷缩维上缠绕着的弦怎么可能测得那半径是1/R呢？这种忧虑是很正常的，不过，问题本身却表述得不够准确。你知道，我们在说弦绕着半径为R的圆时，必然利用了某个距离定义（这样“半径为R”才有意义）。但这一个距离定义却是与未缠绕的弦模式相关的——即与振动模式相关。从这个定义——也只有从这个定义——看，缠绕的弦在空间的卷缩维展开。然而，从第二个距离定义——即与缠绕弦相关的那个定义——看，它们却是局限在空间的一点，就像第一种定义观点下的振动弦一样，而那“一点空间”的半径在它看来是1/R，如正文所讲的。

这多少说明了缠绕和未缠绕的弦所测得的半径是互为倒数的，但是，这一点还是有点儿难以捉摸，看来我们应该为对数学感兴趣的读者说说它背后的数学。在普通的点粒子量子力学里，距离与动量（本质上还是能量）通过富里叶变换相联系。就是说，在半径为R的圆周上的位置本征态｜x＞可以定义为｜x＞ $=\Sigma_N e^{ixp}$｜p＞，这里p＝v/R，而｜p＞是动量本征态（类似于我们所说的弦的均匀振动模式——没有形变的整体运动模式）。但在弦理论中，还存在另一个位置本征态的概念，$\tilde{X}$＞，通过缠绕弦的状态来定义：$\tilde{X}$＞ ，这里$\tilde{p}$＞是缠绕弦的本征态，$\tilde{p}$＝wR。根据这些定义，我们马上发现，x以2πR为周期，$\tilde{X}$以2π/R为周期。这说明x是半径为R的圆周上的位置坐标，$\tilde{X}$是半径为1/R的圆周上的位置坐标。说得再具体些，我们现在可以让两个波包｜x＞和$\tilde{X}$＞从原点开始随时间演化，从而实现我们的两个操作方法来定义距离。不论用哪种方法，圆周的半径都正比于波包四到原来状态所需的时间。由于能量为E的状态伴着相因子Et演化，所以对振动模式来说，时间（从而也是半径）为t～1/E～R；而对缠绕模式来说，t～1/E～1/R。

5.对数学感兴趣的读者可以看到，更准确地说，弦振动的旋数等于卡－丘空间欧拉特证数的一半，这在上一章注3里已经说过了。这个数由$h_{2,1}$与$h_{1,1}$之差猱的绝对值来确定。这里，$h_{p,q}$是（p，q）Hodge数。这两个量分别给出了非平凡同调了一圆（“三维孔”）和同调2－圆（“二维孔”）的数目（精确到一个数值变换）。因此，我们在正文里讲孔的总数，而准确地说，族数依赖于奇数维和偶数维孔洞数之差的绝对值。然而结果是相同的。例如，如果两个卡－丘空间的差别在于各自的$h_{2,1}$和$h_{1,1}$Hodge数是相互交换的，粒子族数———以及“孔”的总数———是不会改变的。

6.这个名字源于这样一个事实：“Hodge钻石”———卡－丘空间中不同维的孔洞的数学概括———对一对卡－丘空间来说是互为镜像

反射的。

7.猿镜猿像猿对猿称这一名词也用于物理学的其他完全不同的场合。如我们在第7、8章讨论过的手征性问题———即宇宙是否是左 - 右对称的———讲的便是另一种镜像对称。

## 第11章　空间结构的破裂

假如你一个劲儿地拉扯一块橡皮膜，它迟早会破裂的。这个简单的事实近年来一直令许多物理学家在想，构成宇宙的空间结构是不是也可能出现这样的事情呢？就是说，空间结构会分裂吗？当然，也许因为我们把橡皮膜的例子太当真了，而被它引向了歧路？

在爱因斯坦的广义相对论看来，答案是否定的，空间结构不会破裂。[1]广义相对论的方程牢牢植根于黎曼几何，我们在前一章讲过，那是分析空间相邻位置距离关系的扭曲的一个数学框架。为了使距离关系有意义，基本的数学形式要求空间背景是光滑的——这是一个有严格数学意义的概念，不过它的寻常意思也能把握某些基本特征：没有褶皱，没有针眼，没有一小块一小块“黏”起来的痕迹，当然也没有破裂。如果空间结构生出这些不规则的东西，广义相对论方程就会崩溃，预示着这样那样的宇宙灾难——那些灾难的结果显然没有出现在我们运转良好的宇宙中。

有想像力的理论家并没有因此停止他们的想象。多年来，他们一直在思考，也许某个超越爱因斯坦经典理论并融合量子物理学的新物理学体系，会证明空间结构可能出现裂痕、破裂和重新组合。实际上，当人们认识到量子物理学能破坏短距离下的涨落时，就有人怀疑裂痕和破碎可能是空间结构的普遍特征。虫洞的概念（对星际旅行着迷的人该熟悉这个词儿）就是从这样的想象中产生出来的。想法很简单。想象一下，假如你是某大公司的总裁（CEO），总部在纽约世界贸易中心一座塔楼的第90层。你还有一家患难与共多年的伙伴公司，

在中心另一座塔楼的第90层。[66]两家公司当然不可能搬迁。为了往来密切方便，你自然会想，在两座塔楼间搭一座天桥，这样员工们就能自由往来而用不着上下90层楼了。

虫洞也起着类似的作用：它是一个桥梁或隧道，为连结宇宙两个区域提供了捷径。拿二维模型来说，宇宙像图11.1的样子。假如你公司的总部设在图11.1（a）的下面那个圆圈处；通过那段U形路径，从宇宙的一头走到另一头，你可以来到上面那个圆圈处的另一个办公室。但是，假如空间结构可以破裂，生成图11.1（b）的孔洞，而孔洞还能生长“触角”，像图11.1（c）那样结合起来，这样，原来两个遥远的区域就通过一座空间桥梁联系起来了。这就是虫洞。你可以看到，虫洞在某些地方像那座世界贸易中心的天桥，但还有点根本的差别：世界贸易中心的天桥穿过一个存在的空间区域——两座塔楼间的空间。而虫洞则生成一个新的空间区域，因为二维空间整个就是图11.1（a）的样子（在我们的二维例子中）。薄膜外的区域只不过说明原来的图是不够充分的，它把U形宇宙描绘成我们更高维宇宙里的一样东西。虫洞生成新空间，从而也开辟了新的空间领域。

宇宙中有虫洞吗？谁也不知道。如果有的话，我们也不知道它们是微观的，还是可能在宇宙的一个巨大区域展开。但是，评价虫洞是真还是假，基本的一点在于决定空间结构是否可能破裂。

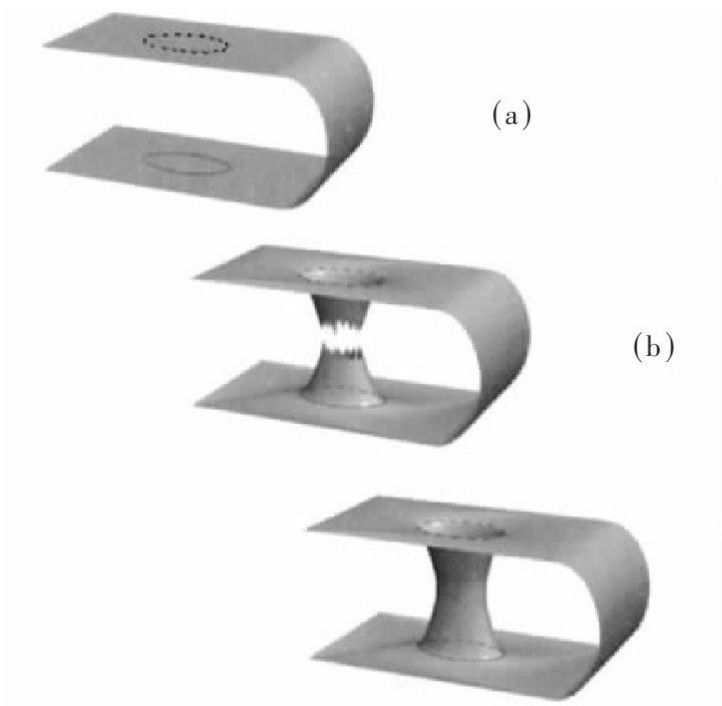



图11.1　（a）在U形宇宙中，从一端到另一端的惟一路径是穿过整个宇宙；（b）空间破裂，虫洞从两端生出；（c）虫洞两端结合，形成一座桥梁——从宇宙一端到另一端的捷径。

黑洞为我们提供了另一个诱人的例子。在这个例子中，空间结构延伸到了极限。在图3.7我们曾看到黑洞巨大的引力场导致了极端的空间卷曲，从而空间结构在黑洞的中心显得破碎了。与虫洞情形不同的是，有许多实验证据支持黑洞的存在，所以关于在黑洞中心发生什么事情的问题，是科学的，而不是幻想的。在这样极端的条件下，广义相对论的方程仍然是失败的。有些物理学家曾提出，破碎的空间结构确实存在着，但黑洞的事件视界（它里面的任何事物都逃不出引力的魔掌）遮住了那个宇宙“奇点”。这个想法使牛津大学的彭罗斯（Roger Penrose）提出一个“宇宙监督假说”，只有在事件视界的遮蔽下才可能出现那种空间奇异性。另一方面，还在弦理论发现之前，就有物理学家猜想，量子力学与广义相对论的恰当结合将证明，那种表

面的空间破裂实际上会被量子行为平滑掉——也可以说，破裂的空间又被“缝合”起来了。

随着弦理论的发现和量子力学与引力论的融合，我们最终会研究这些问题的。尽管现在弦理论还不能完全回答它们，但在过去几年里有些密切相关的问题已经解决了。这一章里我们将讨论弦理论如何第一次确定性地证明在某些物理背景下——在某些方面不同于黑洞和虫洞——空间结构是可能破裂的。

诱人的翻转

1987年，丘成桐和他的学生田刚（现在在麻省理工学院）做了一次有趣的数学考察。他们发现，一定的卡-丘空间形态可以通过我们熟悉的数学步骤变换成其他形态：空间表面破裂，生成孔，然后照精确的数学形式将孔缝合起来。[2]简单地说，他们“黏合”了处于原来那个卡-丘空间内部的一类特殊的二维球面——如皮球的表面，如图11.2。（皮球跟所有普通物体一样是三维的，不过，我们这里只谈它的表面，而不管它的组成材料的厚薄，也不管它所包围的内部空间。皮球表面上的点的位置可以用两个数——“经度”和“纬度”——来确定，因而它的表面跟我们前面讨论的水管的表面一样，是二维的。）然后，他们考虑球面像图11.3那样逐渐收缩成一个点。这幅图和本章后面的图都把卡-丘空间简化了，只突出了关系最密切的那一“小块”，但在头脑中我们应该清楚，这样的形变发生在更大的如图11.2的卡-丘空间。最后，田和丘想象，在尖点处将卡-丘空间轻轻分裂，张开缺口（图11.4（a）），重新黏合起来（图11.4（b）），然后让它膨胀成圆球的形状（图11.4（c）、图11.4（d））。

数学家称这样一个操作序列是一种翻转变换（flop-transition）。那是说，原来的皮球似乎在整个卡-丘空间里“翻转”到一个新的方向。丘、田和其他研究者还注意到，在一定条件下，翻转生成的新卡-丘空间（如图11.4（d））与原来的卡-丘空间（如图

11.3（a））在拓扑学上是不相同的。这个奇特的说法实际上等于说，绝对不可能不经过空间结构的破裂而将初始的图11.3（a）的卡-丘空间变形成为最后的图11.4（d）的卡-丘空间。

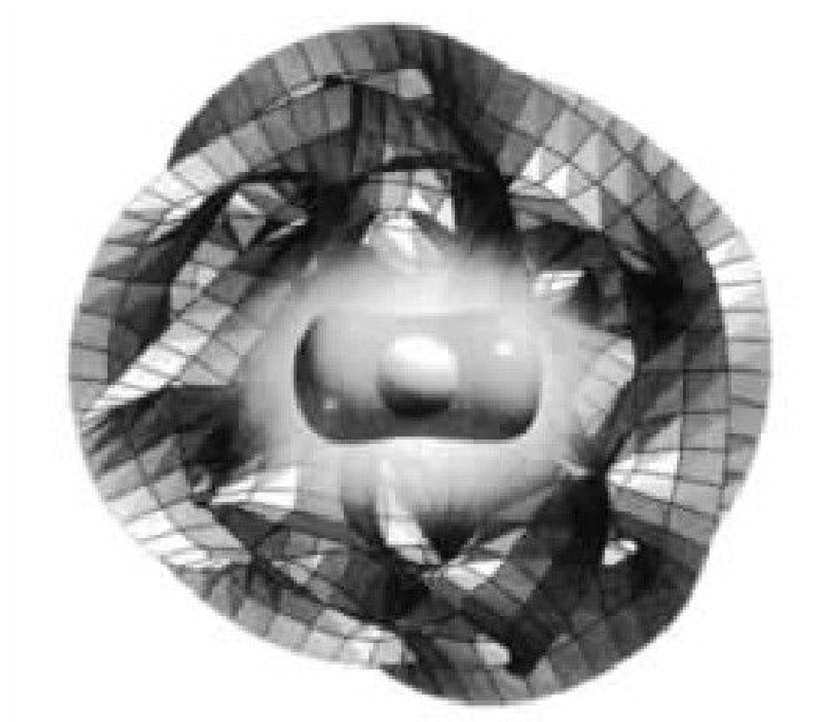
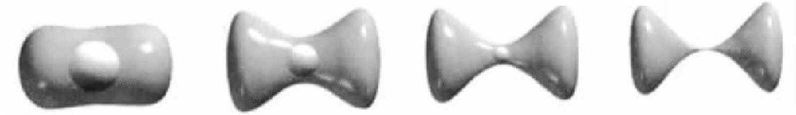


图11.2　在卡-丘空间内部包含着一个球面，特别突出了球所在的区域。



(a)　(b)　(c)　(d)

图11.3　卡-丘空间里的球收缩成一点，使空间结构破裂。在这里和后面的图中，我们简化了卡-丘空间，只画出了有关的部分。

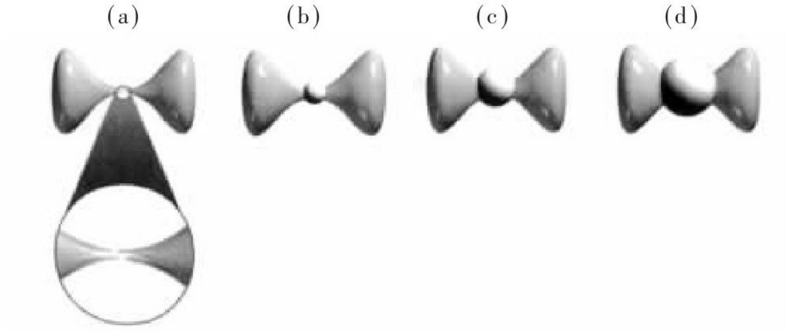


图11.4　破裂的卡-丘空间在尖点处生成一个球面，使表面重新光滑。图11.3中的球被“翻转”过来了。

从数学观点看，丘-田过程的意义在于提供了一个从已知卡-丘空间生成新空间的途径。不过，它的真正潜力还在物理学方面，它提出一个诱人的问题：除了抽象的数学程序外，从图11.3（a）到图11.4（d）的序列真能在自然界出现吗？也许，空间结构果然与爱因斯坦的想象不同，它可能分裂然后像上面讲的那样重新修补好？

镜像图景

自1987年的发现以来的几年，丘常鼓励我去考虑翻转变换是否能在物理学中实现。我没有去想这个问题。在我看来，翻转变换只不过是抽象的数学过程，与弦理论的物理毫不相干。实际上，我们在第10章的讨论中发现卷缩的空间维有一个极小半径，可能有人因此认为弦理论不允许图11.3的球面收缩成一个点。不过，请记住，我们在第10章还讲过，假如是一块空间在坍缩——在这里是卡-丘空间的一个球面——而不是整个维在坍缩，则关于大小半径相同的论证就不适用了。但是，不管怎么说，即使我们不能因为这一点理直气壮地排除翻转变形，空间结构看来仍然不太可能会发生破裂。

可是后来，在1991年，挪威物理学家吕特肯（Andy　Lütken）和阿斯平沃尔（Paul　Aspinwall，我的研究生同学，从牛津来的，现在

是杜克大学教授）提出了一个后来证明是很有趣的问题：假如我们宇宙的卡-丘空间结构会经历空间破裂的翻转变换，那么从镜像的卡-丘空间来看，它会是什么样子呢？为明白提出这个问题的动机，我们需要回想一下，一对镜像卡-丘空间（当然指的是被选作额外维度的那些形式）生成的物理学是相同的，但物理学家为了认识物理而在两个空间遇到的数学困难却大不相同。阿斯平沃尔和吕特肯猜想，从图11.3到图11.4的复杂的数学翻转变换可能有一种简单得多的镜像描述——能更清楚地透视相关的物理图景。

那时候，镜像对称的认识深度还不能回答他们提出的问题。不过，阿斯平沃尔和吕特肯发现，在镜像图景中似乎不会出现翻转变换带来的灾难性的物理结果。大约同时，普里泽和我为寻找卡-丘形态的镜像对的工作（见第10章）也意外将我们引到翻转变换的问题上来。在数学上大家都熟悉，像图10.4那样黏合空间的不同点——我们曾用这个程序来构造镜像对——会产生与图11.3和图11.4中的破裂与缝合相同的几何状态。然而，普里泽和我却没有发现有什么相关的物理学灾难。而且，在阿斯平沃尔和吕特肯的发现（还有他们和罗斯以前的一篇论文）激励下，普里泽和我发现，在数学上我们可以用两种不同的方法来修补空间的破裂。一种方法得到图11.3（a）的卡-丘形态，另一种方法则得到图11.4（d）的形态。这就说明，从图11.3（a）向图11.4（d）的演化在大自然是能够发生的。

到1991年底，至少有几位弦理论家强烈感到，空间结构能发生破裂，但还没有人掌握能确定或否定这种惊人的可能性的数学工具。

一步步往前

1992年，普里泽和我断断续续地努力证明过空间结构能发生空间破裂的翻转变换。我们的计算得出些零星的间接证据，但还没找到确定的证明。那年春天，普里泽去访问普林斯顿高等研究院，把我们最近关于在弦理论的物理条件下空间破裂翻转变换的一些认识私下告诉

了惠藤。普里泽大概讲了我们的想法，然后等惠藤回答。惠藤把头从黑板转过来，两眼望着办公室的窗外。大约过了一两分钟，他才转过头来，告诉普里泽说，如果我们的想法行得通，“那将是很惊人的。”这又激发起我们的热情。可是不久，由于没什么进展，我们两个都去做弦理论的其他课题了。

尽管这样，我还是在思考翻转变换的可能性。几个月过去了，我越来越相信那应该是弦理论的一个不可分割的部分。普里泽和我的初步计算以及我们与莫里森（David Morrison，杜克大学的数学家）富有启发的讨论，似乎都说明惟有这才是镜像对称的自然结果。实际上，在访问杜克期间，莫里森和我在卡茨（Shel-don Katz，来自俄克拉荷马州立大学，那时也在杜克访问）的一些发现的启发下，初步提出了一个证明翻转变换能在弦理论中出现的策略。但当我们坐下来计算时，才发现那是非常艰难的。即使全世界最快的计算机，也需要一百多年才能完成那些计算。我们取得了一点进展，但显然还需要能大大提高我们计算效率的新思想。碰巧，埃森大学的数学家巴提列夫（Victor Batyrev）在1992年春夏的两篇论文无意间揭示了那个思想。

巴提列夫早就对镜像对称感兴趣，特别当坎德拉斯和他的合作者们用它成功解决了第10章最后讲的数球问题以后。不过，他凭一个数学家的眼光，为普里泽和我借以寻找卡-丘空间对的方法感到不安。虽然我们用的工具是弦理论家都熟悉的，巴提列夫后来却告诉我，我们的论文在他看来像“黑色魔术”。这反映了物理学与数学两个学科间巨大的文化差异；当弦理论在模糊它们的界限时，这些差异在两个领域的语言、方法和风格上表现得更显著了。物理学家喜欢先锋派的作风，在寻求问题的解决方法时宁愿改变传统法则，超越大家公认的界线。数学家更喜欢古典风格，习惯按部就班做事情，在前一步没有严格确立以前不会果敢地迈出下一步。两种作风各有优点和缺点；都展开了一条独特的通往创造性发现的道路。两条道路也跟现代与古典音

乐一样，不能讲谁对谁错——一个人选择什么样的方法路线，主要凭他个人的兴趣和修养。

巴提列夫开始在更传统的数学框架下重建镜像流形，他成功了。在台湾数学家阮希石（Shi-Shyr Roan）以前工作的激发下，他找到一个系统地生成镜像卡-丘空间对的数学程序。他的重建程序可以约化为普里泽和我在我们考虑过的例子中发现的程序，但展现了一个以数学家更熟悉的方式表达的更为普遍的框架。

巴提列夫论文的另一方面是多数物理学家以前没有遇到过的数学东西。就我来讲，虽然能把握他论证的要点，却很难理解许多关键的细节。但有一点是清楚的：如果正确理解和应用他文章里的方法，很可能会走出一条认识空间破裂的翻转变换的新思路。

在这些发现的激励下，那年夏天快结束的时候，我觉得自己应该全身心地回到翻转问题上来，莫里森告诉我，他要离开杜克到高等研究院去一年，我还知道阿斯平沃尔也将去那儿做博士后。通过几个电话和电子邮件，我也决定离开康奈尔大学，到普林斯顿去度过1992年的秋天。

策略

要长时间紧张地集中精力做件事情，恐怕很难找到比高等研究院更理想的地方了。它于1930年建在一片如诗一般的森林边的小山坡上，离普林斯顿大学校园只有几英里。人们都说在研究院工作不会受到干扰，当然啦，因为这里本来就没有什么干扰。

1933年，爱因斯坦离开德国以后就来到研究院，在这里度过他的余生。在这幽静、孤独的苦行僧生活的环境里，一位老人在思索他的统一场理论，这是怎样的图景，是不难想象的。这里的空气仿佛也总是弥漫着深沉的思想，它可能令你兴奋，也可能让你感到压抑——这

得看你当时的思想状况是什么样的。

到研究院不久，保尔·阿斯平沃尔和我有一天走在纳索街头（普林斯顿小城的主要商业街），想找一家我们都喜欢的地方晚餐。这可不大容易，因为保尔爱吃肉，而我是个素食者。我们一边走，一边谈自己的生活。谈话中，他问我有没有什么可以做的新东西。我告诉他，是有点儿新东西。然后，我向他详细讲了我觉得重要的事情是应该证明，宇宙如果真是弦理论描绘的那样，则它会发生空间破裂的翻转变换。我还简单讲了我正在探寻的路线，并告诉他，我从巴提列夫的工作看到了新的希望，它大概能弥补我们失去的一些东西。我想这些东西保尔应该是知道的，会为它的前景感到兴奋。然而他没有。现在想来，他那时沉默的原因主要是我们在思想上已经友好地竞争了很久，我们对对方的观点总是有点儿吹毛求疵的。过些日子以后，他转变了看法，我们都全心全意来关注空间翻转问题。

那时，莫里森也来了。我们三个就在研究院的休息室里草拟研究计划。我们都认为，中心目标是要明确从图11.3（a）到图11.4（d）的演化是否能在我们的宇宙发生。但直接攻克这个问题是不可能的，因为描写演化的方程太难了，特别是在空间发生破裂时，更加困难。我们选择了另一种方法，用镜像的图景重新表达这个问题，希望其中的方程会更容易把握一些。图11.5大概说明了这个过程。上面的一行是原来从图11.3（a）到图11.4（d）的演化序列，下一行是同一演化在镜像卡-丘空间里的表现。正如我们很多人已经认识的，它说明在镜像空间里弦理论表现出良好的特性，没有出现灾难性的结果。你可以看到，在图11.5的下面一行里似乎并没有什么破裂。不过，这里出现的真正问题是：我们是不是把镜像对称推到了它的适用范围以外？尽管图11.5上下两行最左端的卡-丘形态能生成相同的物理，但是，在向右端演化的每一步——在中间必然经过破裂和修复的过程——都能让原来的和镜像观点下的物理性质一样吗？

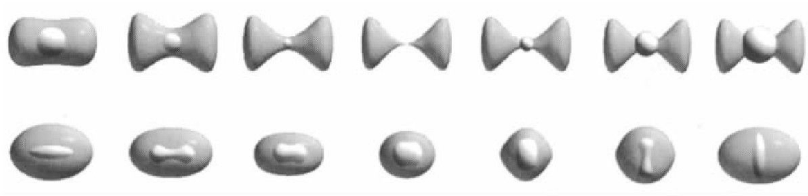


图11.5　一个空间破裂翻转变换（上一行）和设想的镜像过程（下一行）。

虽然我们有很牢固的根据来相信镜像关系对图11.5上面一行所引起卡-丘空间破裂的序列是成立的，但我们也发现，谁也不知道在破裂发生以后上下两行是否还能继续互为镜像。这是一个关键问题。如果它们是镜像的，则镜像空间不会出现灾难就意味着原来的空间也没有灾难，这样我们就证明了弦理论里的空间能发生破裂。我们发现，这个问题可以归结为一种计算：计算原来的卡-丘空间在破裂以后（即图11.5上一行右端的卡-丘形态）的物理性质以及相应的镜像空间（即图11.5下一行右端的卡-丘形式）的物理性质，看它们是否相同。

阿斯平沃尔、莫里森和我在1992年的秋天所做的，就是这个计算。

爱因斯坦暮年归宿的深夜的课堂

惠藤剃刀般的智慧多藏在温和的言谈中，而他的语言常常露着几乎刺人的锋芒。很多人认为，在当今的大物理学家行列里，他是活着的爱因斯坦。甚至还有人说他是有史以来最伟大的物理学家。他对尖锐的物理学问题有永不厌倦的渴求，对决定弦理论的发展方向有着巨大的影响。

惠藤的创造力是源源不断的，还有些传奇的故事。他的夫人娜菲（Chiara　Nappi）也是研究院的物理学家，曾向我们描绘了一个坐在餐桌旁的惠藤：他常常神游到弦理论的边缘，只是需要拿纸和笔计算一些令人困惑的细节时，才偶尔回到现实中来。67另一个故事是听一位博士后讲的。某个夏天，他正好在惠藤隔壁的办公室。他说，当他

痛苦艰难地在桌旁与复杂的弦理论计算搏斗时，常听到有节奏的键盘声不断从惠藤那儿传来，感觉一行行拓荒的文字正从人脑汩汩地流进电脑。

大约一个星期后，我来了。惠藤和我在研究院的园子里聊天，他问我有什么研究计划。我告诉他有关空间破裂翻转的事情和我们正在考虑的证明计划。听到这些想法，他的眼睛亮了，不过，他担心计算会很可怕。他还指出我们计划里的一个薄弱环节，与我几年前与瓦法和瓦纳做过的一项研究有关。但后来发现，他提出的问题只是碰到了翻转问题的边缘，不过这使他开始思考最终的相关而互补的问题应该是怎样的。

阿斯平沃尔、莫里森和我决定把计算分解成两个部分。最自然的分解大概是这样的：先揭示出与图11.5上面一行最后一个卡-丘形态相关的物理，然后对下一行的最后一个卡-丘形态做同样的事情。如果镜像关系没有因为上面卡-丘空间的破裂而破坏，则这最后两个卡-丘空间将跟它们演化之初的两个空间一样，生成同样的物理。（这样表达的问题，避免了卡-丘空间破裂时的复杂计算。）然而，结果表明，计算与上一行最后一个卡-丘形态相关的物理是直截了当的事情，这个方案真正的困难在于，首先确定下一行最后一个卡-丘空间——我们假想的上面那个卡-丘空间的镜像——的准确形式，然后再发现与它相关的物理。

为实现后面这一步——在下一行最后那个空间形态确定的条件下，揭示相关的物理特征——坎德拉斯在几年前就发现了一个方法。不过，他的方法算起来太艰难了，在我们的具体例子中还需要一个更好的计算程序。阿斯平沃尔不但是有名的物理学家，也是一流的程序专家，编程序的任务自然落在他身上。莫里森和我则开始做计划的第一步：弄清那个候选镜像卡-丘空间的准确形式。

就在这个时候，我们觉得巴提列夫的工作能为我们提供一些重要

线索。然而，数学与物理学之间的文化差异——这回是莫里森和我之间的差异——又阻碍了我们的进步。我们需要将两个领域的力量集中起来，去发现图11.5下面那个卡-丘空间的数学形式——如果自然图景中确实可能发生空间破裂，它应该与图11.5上面那个卡-丘空间生成相同的物理。但是，我们两个对对方的语言都还没熟悉到能看清如何达到目标的地步。显然，我们需要补课，需要赶紧走进对方的专业领域。于是，我们决定白天尽可能做计算，晚上上课，既做教授，也当学生：我给莫里森讲一两个小时的物理；然后他给我讲一两个小时的数学。我们经常到夜里11点才下课。

我们日复一日地投入到计划里。进展很慢，但我们能感觉到有些东西就要出现了。这时候，惠藤在加强他以前发现的薄弱环节，取得了重大进展。他的研究是建立一种新的更有力的方法来联结弦理论的物理与卡-丘空间的数学。阿斯平沃尔、莫里森和我几乎每天都跟惠藤坐到一起，他会向我们说明根据他的方法得到的新发现。几个星期过去了，我们逐渐发现，他从完全不同的观点进行的研究竟出人意料地和我们的翻转变换问题走到一起来了。我们觉得，如果不快点儿完成计算，惠藤就会赶到前头去了。

周末的六箱啤酒

对物理学家来讲，友好的竞争是最能让人精神集中的。阿斯平沃尔、莫里森和我，3个人的大脑都在高速运转着。有意思的是，这在莫里森和我是一样的，而阿斯平沃尔则是另一回事了。他身上奇特地体现着英国绅士的个性特征，而且很少玩笑，这大概是他在牛津过了10年学生和研究生的生活留下的印迹。从工作习惯说，他也许是我所见过的最洒脱的物理学家。我们很多人都要工作到深夜，而他的工作从来不超过下午5点。我们周末也工作，而他不会。对他来说，发条拧得太紧，会转得更慢。

到12月初，莫里森和我互相讲课已经几个月，开始有了一点儿回

报。我们离认识要找的卡-丘空间的准确形式已经很近了。另外，阿斯平沃尔的计算程序也刚完成，他等着我们的结果，那是他程序所需要的输入条件。一个星期二的晚上，莫里森和我终于相信我们知道如何识别我们需要的卡-丘空间。那也归结为一个用很简单的计算程序就能完成的过程。星期五下午我们把程序写出来调试，到后半夜，结果出来了。

可那是星期五，下午5点以后的事情。阿斯平沃尔已经回家了，要星期一才回来。没有他的计算程序我们什么事也做不了。莫里森和我真不知道整个周末该怎么过。宇宙结构的空间破裂问题想了那么多年，现在我们已经走到答案的边缘了，怎么还能等下去呢。我们给阿斯平沃尔家里打去电话，让他第二天一早就回来。他开始不愿意，后来还是嘟囔着答应了，不过要我们给他买六箱啤酒，我们答应了。

真理时刻

我们如约在星期六的早上聚在一起。那是一个阳光明媚的早晨，我们玩笑着，气氛很轻松。我说，我一半是想阿斯平沃尔别来，如果来了，我会用15分钟来赞美这个让他第一次走进办公室的周末。他说，保证不会有下一次了。

在我和莫里森共用的办公室里，我们围在莫里森的计算机旁。阿斯平沃尔告诉莫里森如何打开他的程序，向我们演示了需要输入东西的准确形式。莫里森把我们前夜得到的结果化为恰当的格式，就这样开始了。

我们进行的特别计算，大概说来是决定一定粒子种类的质量——也就是，弦在我们花了整整一个秋天来认识的卡-丘空间所在的宇宙中运行时，一定振动模式所对应的质量。依照原来的策略，我们希望这个质量应该与空间破裂翻转生成的卡-丘形态的计算结果一致。后面这个计算相对更容易一些，我们以前已经做过了，结果在我们用的特殊

单位下是3。因为现在做的是可能的镜像数值计算，我们希望得到很接近3但不是3的结果，如3.000001或2.999999，微小的误差来自四舍五入。

莫里森坐在计算机旁，手指在“enter”键上，轻轻一按，他说“开始”，就让程序运行起来。几秒钟后，计算机回到了答案：8.999999。我的心一沉，难道空间的破裂翻转破坏了镜像关系？它们不可能真的发生？不过，我们几乎马上意识到一定出了什么可笑的事情。假如两个空间形式的物理学真不一样，计算机不可能得出一个那么接近整数的结果。假如我们的思想错了，就没有理由期待除随机的数字以外还能有什么别的东西。我们得到一个错误的结果，但它却提醒我们，也许我们是犯了某个简单的算术错误。阿斯平沃尔和我来到黑板前，没多久就发现我们错哪儿了：在一个星期以前做的“简单”计算里，我们忽略了一个因子3，正确结果应该是9。于是，计算机的结果正好是我们想要的。

当然，这种“事后的一致”只能从边缘增强我们的信心。如果我们知道想要的答案，通常很容易找到办法来得到它。我们还需要做别的计算。必要的程序都编好了，做起来也不难。我们在原来的卡-丘形式上计算了另一种粒子的质量，这次十分小心，不会有错了。答案是12。然后，我们又在计算机旁忙开了。几秒钟后，结果出来了：11.999999，是一致的。我们这就证明了假想的镜像空间的确是镜像的，从而空间破裂翻转变换是弦理论物理的一部分。

这时，我一下子从椅子上跳起来，疯狂似地在办公室里跑了一圈。莫里森也笑嘻嘻地坐在计算机旁。不过，阿斯平沃尔的反应却不一样。“那太好了，但我知道会成功的，”他平静地说，“可啤酒在哪儿？”

惠藤的方法

那个星期一，我们满怀胜利地走向惠藤，告诉他我们成功了。他很高兴听到我们的结果。实际上，他也刚找到一个办法来证明发生在弦理论里的翻转变换。他的论证和我们的迥然不同，而且特别说明了为什么这种空间破裂不会产生灾难性后果的微观原因。

他的方法暴露了空间破裂时点粒子理论和弦理论间的差异。关键的一点差异是，在破裂处弦有两种运动形式，而点粒子只有一种。就是说，弦可以像点粒子那样走近破裂，也可以像图11.6画的那样包围着破裂而经过它。总之，惠藤的分析表明，围绕着破裂点的弦——一种不可能在点粒子理论中出现的东西——使周围的宇宙避免了灾难的结果；如果没有它，灾难是一定会发生的。看来，弦的世界叶——回想一下第6章，它是弦扫过空间形成的己维曲面——仿佛提供了一个保护的屏障，消除了空间结构的几何退化所产生的可怕影响。

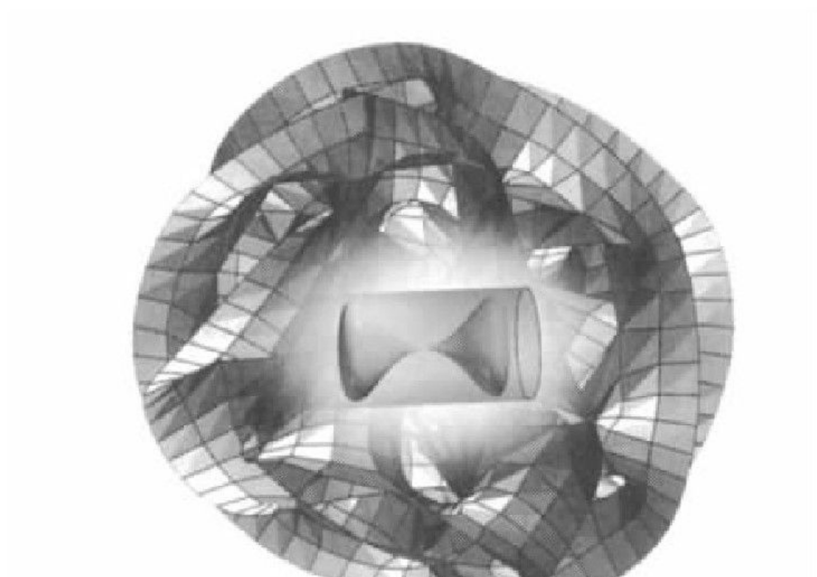


图11.6　弦扫过的世界叶面像一道屏障，消除了与空间结构破裂相关的可能的灾难性影响。

你很可能要问，如果破裂发生的地方没有弦，结果会怎样呢？而且，你还可能想，在破裂发生的那一瞬间，一根弦——一根无限细的线圈——不过像你身上的一根呼啦圈，能遮挡飞来的一群子弹吗？这两个问题的解答在于第4章讨论过的量子力学的一个基本特征。我们在那儿看到，在量子力学的费曼形式里，一个物体，不论是粒子还是弦，都是“摸索着”所有可能的路径从一个地方运动到另一个地方。我们看到的运动是所有可能的组合，每一可能路径在组合中的多少完全决定于量子力学的数学。假如空间出现破裂，则弦可能的运动路径就是图11.6中那些包围破裂点的路径。即使破裂发生时附近没有弦，量子力学考虑的是所有可能弦路径的物理效应，其中就有许多（实际上是无限多）包围破裂点的保护路径。惠藤向我们揭示的就是这些东西，它们消除了可能出现的宇宙灾难。

1993年1月，惠藤和我们三个同时在互联网上发布了我们的论点，通过这种途径，物理学论文可以迅速传遍世界。两篇文章从截然不同的观点描述了所谓拓扑变化转换的第一个例子——那是我们发现的空间破裂过程的专用名词。空间结构是否能发生破裂的老问题就这样由弦理论定量地解决了。

影响

我们已经很好认识了空间能发生破裂而不产生物理学灾难。但是，空间破裂时会发生什么事情呢？会带来什么看得见的影响呢？我们已经看到，周围世界的许多性质都取决于卷缩维的详细结构。于是，你可能认为像图11.5那样神奇的卡-丘空间变换会产生巨大的物理学影响。然而，实际上我们用以描绘空间的二维图像使得那变换看起来比实际发生的更加复杂了。如果能看见六维的几何，我们会发现，空间确实破裂了，但那变化方式是非常“温和”的，像绒毛上的小蛀洞，而不是牛仔裤膝盖上的大口子。

我们和惠藤的结果都说明，像弦振动的族和每一组的粒子类型的

数目这样一些物理特征都不受那些过程的影响。当卡-丘空间通过破裂而演化时，影响的只是每个粒子的质量大小——即弦的可能振动模式的能量。我们的文章表明，这些质量将随卡-丘空间几何形态的改变而连续变化，有的增大，有的减小。然而，最重要的是，当空间破裂出现时，变化的质量并不会出现灾难性的跳跃、尖峰或其他异常行为。从物理的观点看，破裂的瞬间没有什么奇特的表现。

这引出两个问题。第一，我们以上关心的是发生在宇宙多余的六维卡-丘空间里的空间结构破裂，这样的破裂在寻常的三维空间也会出现吗？几乎可以肯定地回答，是的。毕竟，空间就是空间，不论它卷缩成卡-丘形态，还是展开成我们在星光灿烂的夜晚所感觉的茫茫宇宙；即使卷缩的维与展开的维之间有多大区别，那多少是人为产生的。尽管我们和惠藤的分析都依赖于卡-丘空间特别的数学性质，但空间能产生破裂的结果一定有着更广泛的适用性。

第二，这种拓扑改变的破裂会发生在今天或者明天吗？会发生在过去吗？会的。基本粒子质量的实验观测表明，它们的值是相当稳定的。但是，如果我们回到大爆炸以来的早期阶段，即使不以弦为基础的理论也假定有一个基本粒子质量随时间改变的重要时期。从弦理论的观点看，这样的时期当然会发生本章讨论的拓扑改变破裂。在离现在更近的时期，基本粒子质量看起来是稳定的，这说明如果宇宙还在经历着拓扑改变的空间破裂，那过程也该是非常缓慢的——从而它对基本粒子质量的影响微小得我们今天的实验还发现不了。值得注意的是，只要条件满足了，今天的宇宙就可能处在空间破裂的过程中。假如过程很慢，我们就不会知道它的发生。没有发现特别惊人的现象，却引起了极大的兴奋，这在物理学中是少有的事情。那样奇异的几何演化没带来看得见的灾难性结果，这让我们看到弦理论在爱因斯坦的期望之外已经走了多远。

注释

1.喜欢数学的读者会发现，我们实际在问，空间的拓扑是否是动态的——即它是否会改变。注意，虽然我们常用动态拓扑改变的语言，实际上我们常常考虑一个时空的单参数族，它的拓扑像一个单参数函数那样改变。从技术上说，这个参数不是时间，但在一定极限下可以基本把它当成时间。

2.喜欢数学的读者应该看到，这个过程，就是将有理曲线“吹落”到卡-丘流形上来，然后利用这样一个事实：在一定条件下，结果生成的奇点，可以通过特别的小技巧来“修复”。

## 第12章　超越弦：寻找M理论

爱因斯坦在寻求统一理论的漫长道路上，心里想的是“上帝[是否]能以不同方式创造宇宙；就是说，逻辑简单性的要求是否还留着自由的空间。”[68]他的这句话以朴素的形式清楚地表达了今天许多物理学家都相信的一个观点：如果大自然有终极理论，那么支持它的某个特别形式的最令人信服的论证，就是它不可能是相反的东西。终极理论之所以应该有那种形式，是因为那是惟一能描述宇宙而又不产生任何内在矛盾或逻辑荒谬的解释框架。这样的理论宣扬事物就是它本来的样子，因为它只能那样。只要有任何一点变化，不论多么小，都将使理论出现那个“本句话是谎言”的悖论——埋下自灭的种子。

为了认识宇宙呈现那样的结构本来是不可避免的，我们还需要走很长的路去把握一些多年的最深层的问题。那些问题突出了一点神秘：谁或者什么使得那看来无限多的选择成了设计我们宇宙的显然要求？我们常说那是“不可避免的”，这就抹去了那些选择，也回答了那些问题。实在说来，“不可避免性”就是没有选择；它宣扬宇宙不可能是另外的样子。我们将在第14章讨论，没有什么事物能保证宇宙会有如此牢固的结构。不过，追求自然律的这种“刚性”总是现代物理学的统一蓝图的一个核心内容。

到20世纪80年代末，物理学家才发觉，弦理论尽管可能提供一幅独特的宇宙图景，但还不够完美。原因有两点。第一，如我们在第7章简单提过的，物理学家发现实际存在着5种不同形式的弦理论。你可能还记得，它们分别是I型、IIA型、IIB型、杂化O（32）型（简称杂化-O）和杂化$E_8 \times E_8$型（简称杂化-E）理论。它们有许多共同的基本特征——如弦振动模式决定可能的质量和力荷，需要一个10维的时空，卷缩的维应该是某种形态的卡-丘空间，等等——因此，在前面的章节里我们没有强调它们的差别。但是，80年代的分析表明它们的确是有差别的。在后面的注释里你可以看到它们的更多性质，不过我们这里知道两点就够了：它们包容超对称性的方式不同；它们具有的振动模式的细节不同。[1]（例如，I型弦理论除了有我们集中讨论过的闭弦而外，还有两端自由的开弦。）这曾令弦理论家感到疑惑，因为尽管我们需要一个真正的最终的统一理论，但涌现出5种可能的形式来，却令每一种都不够理直气壮了。

第二点偏离“不可避免”的事情更难懂一些。为完全明白这一点，我们应该认识到所有物理学理论都包含着两个部分。一部分是理论的基本思想，通常由数学方程表达；另一部分则由这些方程的解组成。一般说来，一些方程有1个而且只有1个解，而另一些方程有多个（也可能很多个）解。（举一个简单例子，方程“2乘以某个数等于10”只有一个解：5。但方程“0乘以某个数等于0”则有无限多个解，因为0乘以任何数都是0。）所以，即使找到由惟一一组方程组成的惟一一个理论，也不一定得到“不可避免的”结果，因为这些方程可能有许多不同的解。80年代末，人们发现弦理论正处在这样的情形。物理学家在研究5个弦理论中的任何一个的方程时，发现它们确实有许多解——例如，额外的维有多种不同的卷缩形式——每一个解都对应一个不同性质的宇宙。虽然多数宇宙都是作为弦理论方程的有效解出现的，但与我们所知的宇宙似乎没有什么关系。

弦理论得不到“不可避免的”结果，这看起来是很不幸的一个基本

特征。但90年代中期以来的研究为我们带来了极大的新希望：这些特征可能只不过是弦理论家们所用的分析方法产生的。简单地说，弦理论方程太复杂了，谁也不知道它们的精确形式。正是这些近似的方程使一个弦理论迥然不同于另一个。也正是这些近似的方程在5种不同的弦理论背景下出现那么多的解，生成那么多没用的宇宙。

1995年（第二次超弦革命开始那年）以来，越来越多的证据表明，精确的方程（其精确形式我们今天还不知道）可以解决这些问题，从而有助于为弦理论带来“不可避免的”结果。实际上，大多数弦理论家都满意地发现，当精确方程建立起来时，它们会证明5种弦理论原本是密切联系的。5个弦理论像海星的5个触角那样，是同一个整体的不同部分，而我们今天正在努力研究那个整体的性质。物理学家现在相信，他们并没有5个不同的理论，而是有一个把5个理论缝合在惟一一个理论框架的理论。当隐藏的关系显露出来时，问题就一目了然；同样，5个弦理论的统一也将为我们从弦理论看宇宙提供新的有力的视点。

为解释这些东西，我们必须认识弦理论的一些最困难、最前沿的发展。我们必须认识弦理论研究中应用的近似方程的本质和内在局限；我们必须熟悉物理学家借以克服某些近似的灵巧办法——那些技术总称对偶性。接下来，我们必须跟着这些技术的逻辑路线去发现上面提到的那些惊人的结果。但你用不着担心，真正困难的事情弦理论家们已经做了，我们只需要解释他们的结果就行了。

不过，我们要讲的有许多看似分离的东西，在这一章里很容易看见了树而失去了森林。所以，如果你什么时候觉得讨论太复杂了，想急着去看黑洞（第13章）和宇宙学（第14章），请你回头来看看下面的一节，它概括了第二次超弦革命的要点。

第二次超弦革命

图12.1和图12.2概括描绘了第二次超弦革命的基本思想。在图12.1中我们看到，在没能超越物理学家用来分析弦理论的传统近似方法以前，是怎样的情形。5个理论看起来是完全分离的。但是，据今天的研究，我们发现那5个弦理论就像图12.2中海星的5只触角那样，是一个包容一切的框架。（实际上，在本章最后我们还会看到第六个理论——海星的“第六只触角”——也将融入这个统一。）这个囊括四方的框架现在暂时叫作M理论，我们下面将明白这是为什么。图12.2是寻求终极理论的一块里程碑。弦理论中看似毫无牵连的研究现在编织成为一个独一无二的统一的理论，那可能就是我们寻求已久的包罗万象的理论。

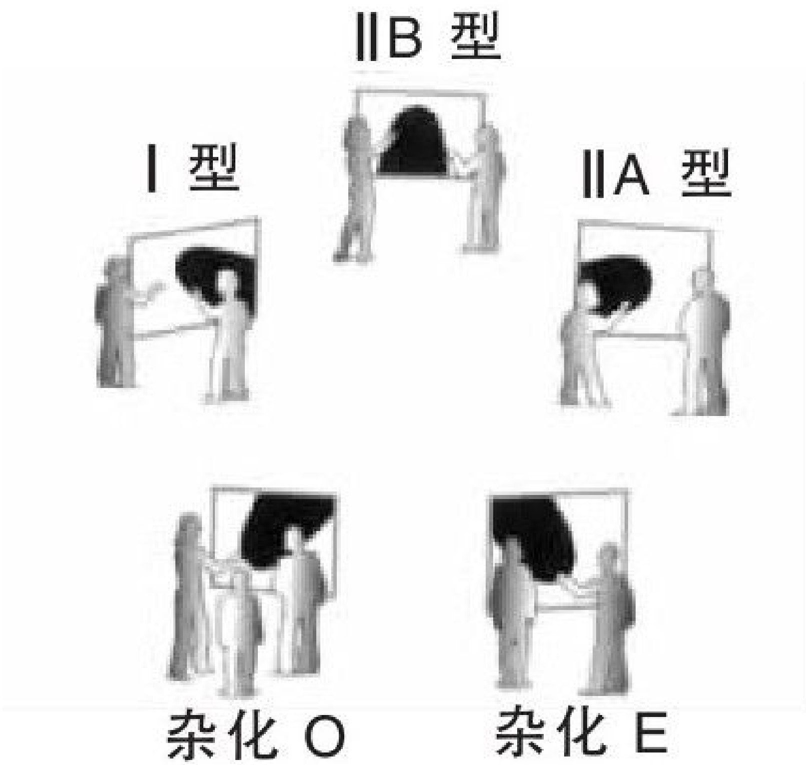


图12.1　多年来，在5个弦理论上做研究的物理学家认为他们是在完全独立的理论上工作。

虽然还有好多事情要做，但物理学家已经发现了M理论的两个基本特征。第一，M理论有11维（10维空间和1维时间）。我们记得，卡鲁扎曾发现多1个空间维会意想不到地将广义相对论与电磁学结合起来；弦理论家也发现，在弦理论中，多1个空间维——在我们前面讨论的9维空间和1维时间之外的1维——会令人满意地将弦理论的5个不同形式综合在一起。而且，这多余的1个空间维并不是凭空生出来的，而是早就存在了。弦理论家现在知道，20世纪七八十年代得到9维空间和1维时间的方法是近似的，精确的计算（现在可以完成了）证明还有1个空间维，我们以前都把它忽略了。

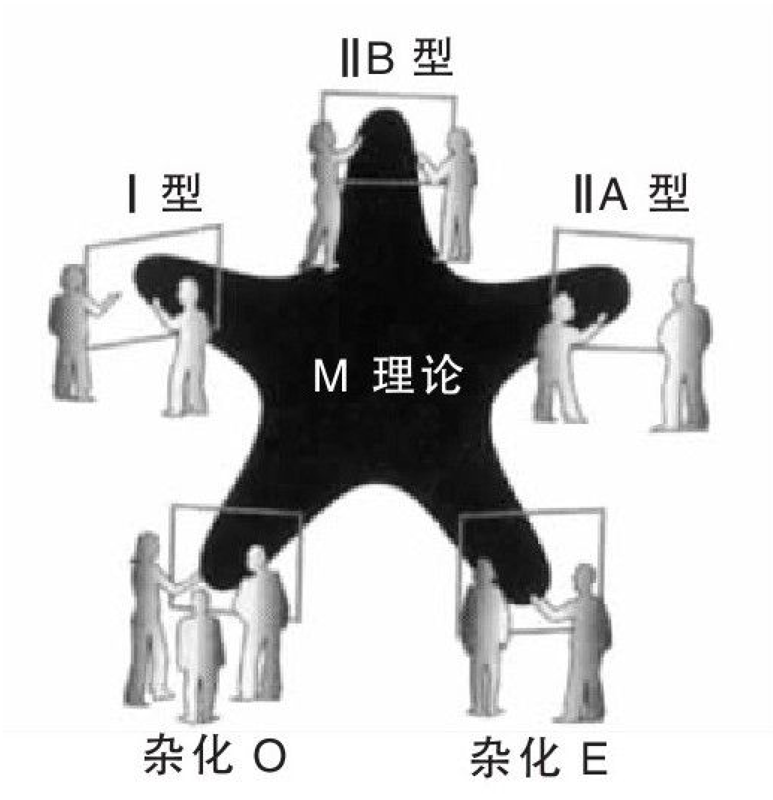



图12.2　第二次超弦革命的结果表明，5个弦理论实际上是一个暂时被称为M-理论的统一框架

的一部分。

我们发现的M理论的第二个特征是，它不仅包含振动弦，还包含着别的东西：振动的2维薄膜、涨落的3维液滴（也叫“3-维”）以及其他一些物质的构成元素。M理论的这些特征也跟11维一样，是我们的计算从90年代以前的近似方法中解脱出来的结果。

除了这两点发现和近几年来的其他一些认识而外，M理论的许多本性的东西仍然是一个个的“谜”——这就是人们说的“M”（在英文是mysterious，在中文是mi）的意思。全世界的物理学家都在以巨大的热情去探求那谜一般的理论，这也成为21世纪物理学的核心问题。

近似方法

物理学家从前用来分析弦理论的方法的局限源于所谓的微扰论。微扰论说的是，对某个问题做一近似处理，得到一个大概的结果，然后更仔细地考虑原先忽略的细节，从而系统地提高近似的程度。在许多科学领域它都起着重要作用，在弦理论的认识中也是基本的方法。现在我们来看，在日常生活里也常能遇到它。

假如某一天你的车出毛病了，你找到一个机械师，请他给检查一下。机械师看过后告诉你一个坏消息：你的车需要换一台新的发动机，一般大约需要900美元。这是很粗略的近似，你希望仔细检查后能得到更细一些的情况。几天以后，机械师告诉你，经过运行检查，你还得换一个调节器，大约50美元。这样，修车的费用更准确了，大约是950美元。最后，你去取车时，他把所有费用加起来，给你一张987.93元的账单。他解释说，那包括950元的发动机和调节器，另外27元是散热器的风扇皮带，10元是电线；最后还有0.93元是绝缘螺栓。原先粗略估计的900元，最后经过一点点的补充，变得准确了。用物理学的语言说，这些一点点的东西都是对原来估计的微扰。

恰当而有效地运用微扰论可以使原来的估计很接近最后的结果；

应用微扰论时，原来忽略的细节不会太大地影响最后的结果。但是，有时候你会发现最后结果与原来的估计差别大得惊人，技术上说这是微扰论的失败，你可能还有更富感情的说法。这说明原来的近似不是最后结果的恰当指南，因为修正的东西不是小小的偏差，而是大大地改变了原来的粗略估计。

在前面的几章我们简单说过，我们关于弦理论的讨论都依赖机械师用的那种微扰方法。我们常说的对弦理论的“不完全认识”，都这样那样地源于这种近似方法。现在，我们在不那么抽象但比机械师离弦理论更近的情形来讨论微扰方法，从而更好地理解它为什么是“不完全”的。

微扰论的一个经典例子

运用微扰论的一个经典例子是认识地球在太阳系中的运动。在这样巨大的距离尺度上，我们只需要考虑引力；但如果不做进一步的近似处理，方程仍然是极端复杂的。我们记得，据牛顿和爱因斯坦的理论，任何事物都对别的事物产生引力作用，这样，自然得到一个在数学上难以应付的复杂的引力“混战”，牵涉到地球、太阳、月亮和其他行星，原则上还包括所有其他的天体。你可以想象，考虑这么多的影响是不可能的，也决定不了地球的准确运动。实际上，即使只有3个天体，方程也会复杂得没人能完全解决它们。[2]

但不管怎么说，我们能用微扰的方法以很高的精度预言地球在太阳系里的运动。与太阳系的其他星体相比，太阳的质量最大；与其他恒星相比，太阳离地球最近。这样，太阳对地球运动的影响远远超过了其他所有的天体。所以，我们可以只考虑太阳的引力作用来获得一个粗略的估计。在许多情况下，这样的估计是足够好的。必要的时候，我们还可以考虑次要的一些天体的引力效应，如月亮和当时经过地球的行星，这样可以使估计更加准确。当引力越来越多时，计算也开始变得困难，但我们还是较清楚微扰论的原则：太阳-地球引力相互

作用为我们近似解释了地球的运动，而其余复杂的引力作用只是对那个解释的一系列越来越小的修正。

微扰方法适用于这个例子的原因在于，这里有一个起支配作用的物理学效应，它的理论描述相对说来更简单。但事情并不总是这样的。例如，假如我们对一个由3颗质量相近的天体组成的三星系统（3颗星相互环绕着运动）感兴趣，就找不出哪个引力关系的影响比别的更大。这样，没有能用来作粗略估计的一个相互作用，而别的效应也不只是一点小小的修正。如果我们硬从两个星体间的引力作用中选一个来运用微扰的方法，用它作一个粗略的估计，我们很快就会发现那是错误的。计算将证明，考虑第三颗星所带来的对原来估计的运动的“修正”不是很小，而是与那粗略的近似一样重要。我们很熟悉这一点：3个人跳霍拉舞一点儿也不像两个人跳探戈。巨大的修正意味着原来的近似离题太远，从而整个计划都不过是一个幻想。我们应该注意，那不单是第三颗星产生的巨大影响的问题，还有更严重的像多米诺骨牌那样的一连串反应：第三颗星极大影响着原来两颗星的运动，而那两颗星反过来也影响着第三颗星的运动，然后它又会影响那两颗，等等。在这个引力作用网中，每一个都同样重要，因而必须同时加以考虑。在这种情况下，我们常常只能靠计算机的神力来模拟可能的运动结果。

这个例子说明，在应用微扰法时，重要的是决定假设的粗略估计是否真是近似的；如果是，那么哪些细节、多少细节还应该考虑进来才能达到需要的精度水平？如我们现在讨论的这几点，对于将微扰工具用于微观世界的物理过程是特别重要的。

弦理论的微扰方法

弦理论里的物理过程建立在振动弦之间的基本相互作用基础上。我们在第6章结束时讲过[69]，那些相互作用包括图6.7的弦圈的分离与结合。为方便起见，我们把图重新画在这里（图12.3）。弦理论家已

经证明了图中示意的过程可以与准确的数学公式联系起来——公式表达了每一根弦对其他弦的运动会产生什么影响。（在细节上，5个弦理论的公式有区别，但现在我们要忽略那些难以把握的特征。）如果没有量子力学，这些公式将是弦相互作用的终点。但是，不确定性原理决定的微观涨落却意味着弦-反弦对（两根振动模式相反的弦）可以在瞬间产生，能量是向宇宙“借”的——不过两根弦必须在足够短的时间里湮灭，然后把能量“还”给宇宙。这样的一个弦对，在量子涨落中生成、靠借来的能量存在，从而必然很快重新结合成一根弦圈，因此被称作虚弦对。虽然它们是瞬间存在的东西，却将影响相互作用的具体性质。

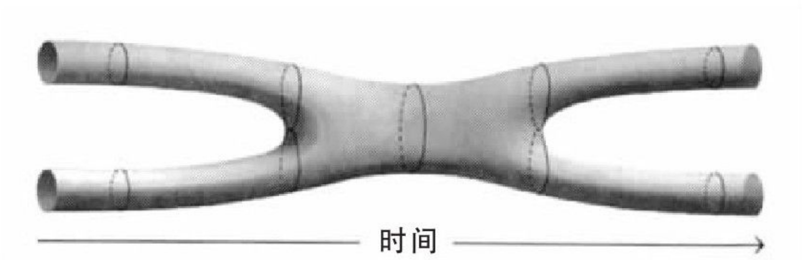



图12.3　弦通过分离和结合发生相互作用。

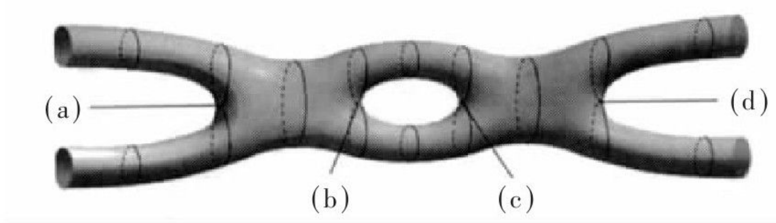



图12.4　量子涨落引发弦-反弦对的生成（b）和湮灭（c），使相互作用更加复杂。

虚弦对如图12.4所示。原来的两根弦“突然”在图中的（a）点相遇，在那里结合成一根弦圈，圈向前运动，在（b）点剧烈的量子涨落生成虚弦对，虚弦对运动到（c）湮灭，又还原成一根弦。最后，

这根弦在（d）点放出能量，分裂成两根弦，沿不同方向运动。图12.4中间有一个环，于是物理学家称它为“1圈”过程。跟图12.3一样，图12.4也联系着一个精确的数学公式，它概括了虚弦对对原来两根弦的运动产生的影响。

不过这个过程还没有结束，因为量子涨落可以引发任意多的瞬间虚弦对，从而生成一个虚弦对的序列。这样便形成圈数越来越多的图，如图12.5。每一个图都为描述有关过程提供了简单适用的方法：两根过来的弦结合成一根弦，量子涨落使它分裂成虚弦对，向前运动，然后湮灭，形成一根弦，在运动中又生成另一虚弦对，如此演进下去。对这些图，每个过程也有对应的数学公式，同样概括了虚弦对的原来两根弦的运动的影响。[3]

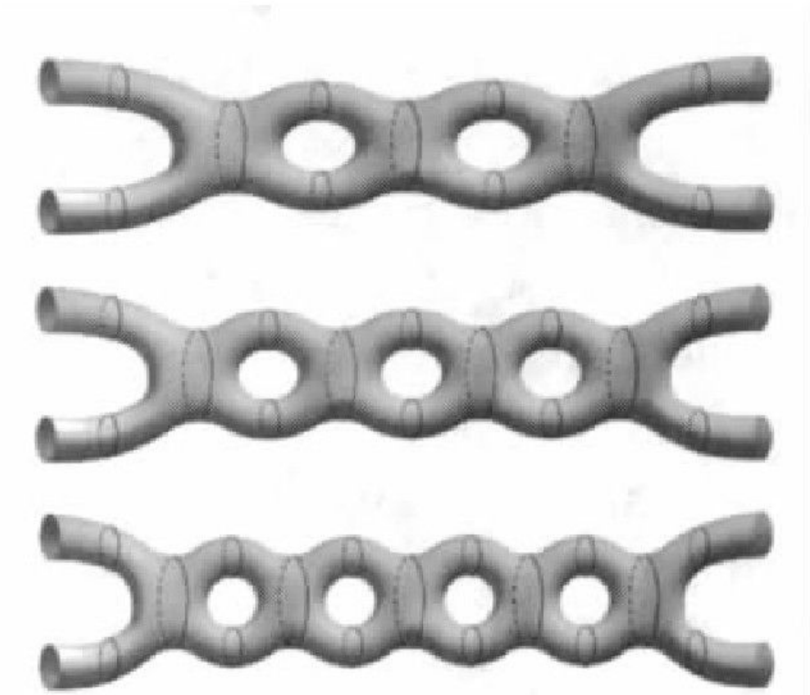


图12.5　量子涨落引起无数次的弦-反弦对的生成和湮灭。

我们在前面看到，你付修车费的时候，机械师在原来估计的900美元外增加了更具体的款项，50元，27元，10元和0.93元；为了更准确认识地球在太阳系的运动，我们在太阳影响之外还考虑了月亮和其他行星的影响。同样，弦理论家证明，两根弦的相互作用可以通过把无圈（没有虚弦对）、1圈（1个虚弦对）、2圈（两个虚弦对）等图的数学表达式加在一起来认识，如图12.6。

为进行精确的计算，我们需要把与圈数越来越多的图相关联的数学表达式加在一起。但是，因为这种图有无限多个，而圈数越多，相关的数学计算也越困难，所以这实际上是不可能的。不过，弦理论家将这些计算转到了微扰论的框架下，这么做的基础在于他们的猜想：零圈过程能得到很好的近似估计，圈图产生一些修正，圈越多，效应越小。

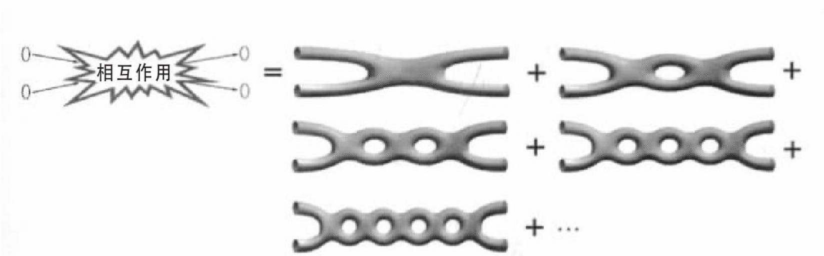



图12.6　一根弦与另一根弦的相互作用的净效应等于各个圈图的影响的总和。

实际上，我们所知的关于弦的几乎所有事实——包括前面章节里讲过的许多东西——都是弦理论家通过用这样的微扰方法进行详尽和精细的计算而发现的。但这些结果是否可信，还要看只考虑图12.6的前几个图而忽略更多圈图的粗略估计是否真是一个近似的估计。这引出我们的一个关键问题：我们的近似真的近似吗？

近似真的近似吗

那要看情况，虽然与圈图相关的数学公式随圈数的增多而变得越来越复杂，弦理论家还是发现了一个基本特征。正如绳子的强度决定

着它是否可能被拉断或者拧断，同样也存在某一个数，确定着量子涨落是否能将一根弦分裂成两根，产生瞬间的虚弦对。这个数就是所谓的弦耦合常数（更准确说，5个弦理论有各自不同的耦合常数，这一点我们马上要讨论）。这个名字说得好：弦耦合常数的大小描述了3根弦（原来的一根和分裂成的两根）的量子涨落的关联有多强——就是说，它们彼此的耦合有多紧。从计算公式看，耦合常数越大，量子涨落越可能使原来的弦发生分裂（然后再结合）；耦合常数越小，虚弦瞬时产生的可能性就越小。

我们很快要讲在任何一个弦理论中决定弦耦合常数的问题，不过，我们凭什么说它是“大”还是“小”呢？这一点，弦理论的数学基础已经证明了，区别“大”与“小”的界线是1。意思是这样的：如果弦耦合常数的值小于1，则数量越多的虚弦对越不可能瞬时产生而存在——就像闪电，在同一地方总不太可能多次出现的；然而，如果耦合常数大于或等于1，则很可能出现越来越多的虚弦对。[4]关键的一点是，如果弦耦合常数小于1，圈图的贡献将随圈数的增多而减小。这正是微扰论方法所需要的，因为它说明即使忽略了除前几个圈图而外的所有过程，也能得到很准确的结果。但是，如果弦耦合常数不比1小，则圈图的贡献将随圈数的增大而增大。这就像三星系统的问题，微扰方法失败了。原来提出的无圈过程的粗略近似这时不近似了。（这里的讨论同样适用于任何一个弦理论——某个理论下的弦耦合常数值决定着微扰近似方法的有效性。）

这将我们引向另一个重要问题：弦耦合常数是多少（或者更准确问，5个弦理论各自的耦合常数是多少）？今天，没人能回答这个问题。这是弦理论的最重要问题之一。我们可以确信，只有耦合常数小于1才可能保证微扰框架下的结果是正确的。而且，弦耦合常数的精确数值将直接影响不同弦振动模式所携带的质量和力荷。这样，我们看到，许多物理性质都依赖于弦耦合常数。因此，我们应该更近地去看看，为什么关于它（在5个弦理论中）的数值的重要问题现在还没

有答案。

弦理论方程

决定弦的相互作用的微扰方法也可以用来决定弦理论的基本方程。大体上说，弦理论的方程决定着弦的相互作用方式，而反过来，弦的相互作用方式也直接决定着弦理论的方程。

一个基本的例子是，在5个弦理论中，各自都有一个用以决定理论的耦合常数的方程。然而，物理学家今天在每一个弦理论中只能用微扰方法估计少数几个相关的弦作用圈图，得到一个近似的方程。近似方程告诉我们的不过是，在5个弦理论的任何一个里，弦耦合常数都有一个这样的数值，它乘以零的结果是零。这太令人失望了；因为任何数乘以零都是零，以任何值作耦合常数都能满足方程。这样，在任何一个弦理论中，关于耦合常数的近似方程等于什么也没说。

这时候，在5个弦理论中还有另一个方程，是提出来决定展开和卷缩的时空维的具体形式的。我们现在有的这个方程的近似形式比关于耦合常数的方程严格得多，但它还是允许有多个解。例如，4个展开的时空维连同卷缩的6维卡-丘空间构成解的一类，但也有其他可能，展开维与卷缩维的数目还可以有另外的划分。[5]

从这些结果我们能得到什么呢？有3种可能。第一，从最悲观的可能说，尽管每个弦理论都有方程来决定耦合常数和时空的维度与几何形式——其他理论不可能回答的问题——但即使这些方程的精确形式（当然我们还不知道），也允许大量的解，从而将根本削弱理论的预言能力。假如真是这样，那就成了一道障碍。因为弦理论承诺自己能够解释宇宙的那些特征，而不是要我们从实验观测去发现它们，然后多少随意地把它们塞进理论。我们在第15章还要回来讨论这个可能。第二，近似弦方程的令人讨厌的随意性可能暗示着在我们的论证中存在微妙的缺陷。我们是在用微扰的方法来决定弦耦合常数的值，

而我们讲过，微扰法只有在耦合常数小于1时才有意义；这样，我们的计算可能就是在未经证明地假定结果本身——即假定计算结果小于1。我们的失败则很可能说明那假定错了，也许5个弦理论的耦合常数都大于1。第三，弦理论那讨厌的随意性可能源自我们用的近似方程。例如，即使某个弦理论的耦合常数小于1，理论的方程也还是可能依赖于所有圈图的贡献。就是说，更多圈图的一点点修正的累积可能会根本改变近似方程——允许有多个解的近似方程——将它改造成更加严格的准确方程。

到20世纪90年代初，多数弦理论家从后两种可能清楚地认识到，理论的进展实在太依赖于微扰论的方法了。他们几乎都认为，下一步的突破需要一种非微扰的方法——它不受近似计算的约束，从而可能远远超越微扰论框架的极限。在1994年的时候，寻找这样的方法似乎还是幻想，但有时幻想也能成为现实。

对偶性

世界各地的几百名弦理论家每年都要聚会一次，总结一年来的成绩，评估各种可能研究方向的优缺点。根据一年的进展情况，人们常常可以预言与会者的兴趣和热情。20世纪80年代中期，在第一次超弦革命的火红年代，这些会总是洋溢着激情和喜悦。物理学家们普遍希望能在短时间内完全认识弦理论，能证明它就是那个宇宙的终极理论。现在想起来，那是太天真了。在后来的年月里，人们发现弦理论有许多深奥的难以捉摸的问题，无疑需要付出长期艰苦的努力才能认识它们。以前那些不切实际的期望曾带来过激情；但当事情没能一下子如愿时，许多研究者就心灰意冷了。20世纪80年代末的弦理论会议就反映了这种理想幻灭后的低落情绪——物理学家带来了有趣的结果，但激不起人们的热情。甚至有人建议这样的年会别再开了。但在20世纪90年代初，情况好起来了。经过不同的突破（有些我们在前面讨论过了），弦理论又恢复了活力，研究者也焕发出乐观的激情。不过，似乎谁也没能预料，1995年3月在南加利福尼亚大学的弦理论年

会上会发生什么事情。

该惠藤讲话的时候了。他走上讲台，发表了一篇点燃第二次超弦革命的演讲。他在杜弗（Duff）、胡尔（Hull）、汤森（Townsend）的早期工作的激发下，在施瓦兹和印度物理学家A·森（Ashoke Sen）等人发现的基础上，提出了一个超越弦理论的微扰认识的纲领。那纲领的核心部分是所谓对偶性的概念。

物理学家们用对偶性来说明那些看起来不同实际上可以证明描写完全相同物理的理论模型。我们来看一个“平凡的”对偶性的例子：实质一样的理论只不过因为表达方式不同而显得不同。如果你只懂中文，那么你可能不会立刻认出用英文写的爱因斯坦的广义相对论。不过，两门语言都精通的物理学家可以很容易把一种语言译成另一种语言，确立二者的等价性。我们说这个例子是“平凡的”，是因为从物理学的观点看，语言的翻译没带来任何东西。如果一个既懂英文也懂中文的人研究广义相对论的一个难题，不论用哪种语言，问题都是一样困难的。沟通两样语言，并不产生任何新的物理认识。

非平凡的对偶性的例子是，同一物理状态的不同描述确实会产生不同和互补的物理学认识与数学分析方法。实际上，我们已经遇到过两个对偶性的例子。在第10章我们曾讨论过，在卷缩维半径为R的宇宙中的弦理论也可以描述为在卷缩维半径为1/R的宇宙的理论。这是两个不同的几何，但因弦理论的性质，它们在物理上是完全相同的。镜像对称是另一个例子。6个额外空间维的两个不同的卡-丘空间——乍看起来迥然不同的两个宇宙——具有完全相同的物理性质。它们为同一个宇宙提供了两个互相对偶的描述。特别重要的是，这里的情形与中英文的对译不同，两个对偶的描述产生了重要的物理发现，如维的极小半径和弦理论中的拓扑变换过程。

惠藤在“95弦”年会上的演讲中提出了一种新的深刻的对偶性的证据。正如我们在这一章开头简单讲的那样，他指出，5个弦理论尽管

看起来有不同的基本结构，但都是同一基本物理学的不同表达方式。于是，我们并不是有5个不同的弦理论，而是有5扇通向同一个基本理论框架的窗口。

90年代中期的弦理论进展之前，像对偶性这样的宏大构思只是物理学家曾经有过的梦想，实际上几乎没人讲出来，因为它太离奇了。如果两个弦理论在结构上大相径庭，人们很难想象它们能是同一基本物理学的不同描述。不过，通过弦理论的神奇力量，越来越多的证据说明5个弦理论确实是对偶的。而且，正如我们将讨论的，惠藤证明可能还有第六个理论走进这个熔炉。

这些思想密切关联着我们在上一节最后讲的关于微扰方法的适用性问题。因为5个弦理论在弱耦合时才表现得各不相同——所谓弱耦合说的是理论的耦合常数小于1。物理学家靠的是微扰方法，所以他们有时不可能回答这样的问题：如果耦合常数大于1，即所谓强耦合的行为，那些弦理论该有什么性质呢？惠藤等人则宣布，这个关键的问题现在可以回答了。他们的结果令人信服地指出，与我们尚未讲过的第六个理论一起，这些弦理论的强耦合行为都有一个对耦的描述，那是另一个理论的弱耦合行为的描述。

为更具体地把握这个思想，我们应该记住下面的例子。有两个与世隔绝的人，一个喜欢冰，奇怪的是他从没见过水（冰的液态形式）；另一个喜欢水，当然，他从没见过冰。一个偶然的机会，两人相遇了。他们决定组队远征沙漠。爱冰者被爱水者的光滑透明的液体迷住了，而爱水者也惊讶地看着爱冰者带的晶莹的固体。两个人都不知道在水与冰之间存在着深层的联系；在他们看来，这是两样全然不同的物质。可是，当他们走进大漠火辣辣的太阳时，才惊奇地发现冰慢慢化成了水；而在大漠寒冷的夜晚，他们同样惊奇地发现液态的水慢慢结成了固态的冰。他们终于认识到，这两种他们原以为毫不相干的物质竟是密切联系的。

5个弦理论间的对偶关系多少有点儿相似：大体上讲，弦耦合常数起着类似于沙漠例子中温度的作用。5个弦理论的任何两个乍看起来都像冰和水那样显得截然不同，但当各自的耦合常数变化时，这些理论却相互转化了。当温度升高时，冰转变成水；同样，在耦合常数增大时，一个弦理论可以转变成另一个。我们经过漫长的征程才发现所有的弦理论都是同一个基本物理结构的对偶描述——就像冰与水，不过都是$H_2O$的具体表现。

这些结论的理由几乎完全依赖于对称性原理的应用。我们下面来讨论这一点。

对称性的力量

多年来，几乎没人想过去研究大耦合常数情况下5个弦理论的任何性质，因为没人知道离开微扰论还能做什么。不过，在20世纪80年代末和90年代初，物理学家已经取得了一些虽然缓慢但是持续的进展，他们认准了某些特别的性质——包括一定的质量和力荷——是某个弦理论强耦合物理的一部分，然而也是我们计算力所能及的。这些显然超越了微扰方法的计算在驱动第二次超弦革命中起着核心作用，而它们的力量来自对称性。

对称性原理为认识物理世界的许多事物提供了洞察的工具。例如我们讲过，物理学定律从来不认为宇宙的某个地方或某一时刻是特别与众不同的，这个古老的信念使我们能够相信，今天的这个地方的定律在其他时刻的其他地方也同样发生作用。这是一个大例子，而对称性原理在不那么宏大的背景下也一样重要。例如，你目睹了一次犯罪，可你只看到了罪犯的右脸；但警察画家可以根据你提供的情况画出罪犯的整张脸。这就是对称性。尽管一个人的左脸和右脸存在一定差别，但基本上还是对称的，一边的脸完全可以用来作另一边的良好的近似。

在广泛的不同领域的应用中，对称性的力量表现在它能以非直接的方式——那通常比直接的方法容易得多——确定事物的性质。当然，为认识仙女星座的基本物理，我们可以到那儿去，寻找一个绕着某颗恒星旋转的行星，在那儿建加速器，做我们在地球上做过的实验。但借助于位置变化下的对称性这一非直接的方法，事情会容易得多。我们也可以直接去追踪那罪犯的左脸的特征，但更简单的办法还是借助脸的左右对称性。[6]

超对称性是一个更抽象的对称性原理，它联系的是具有不同自旋的基本物质组成的物理性质。从实验结果看，至多只有些零星线索表明微观世界里有这种对称性，但根据我们以前讲过的理由，可以相信它确实是存在的。超对称性当然是弦理论的一个组成部分。20世纪90年代，在高等研究院塞伯（Nathan Seiberg）的开拓性研究的指引下，物理学家发现超对称性像一把利剑，能以非直接的方式解决某些重要的纷纭复杂的难题。

即使不了解理论错综复杂的细节，如果知道它有超对称性，我们也能给它所具有的性质提出严格的约束。举一个语言的例子。如果有人告诉我们在一张纸条上写着一串字母，其中“y”出现过3次；纸条封在一个信封里。如果没有别的消息，我们无法猜测这个字母序列——我们所知道的只是它可能是一个完全随机的有3个“y”的序列，像mocfojziyxidqfqzyycdi，或者任何其他序列，有无限多的可能。这时，又有人告诉我们两条线索：那张纸条写的是一个英文单词，而且，在所有含3个“y”的单词中，它是字母最少的一个。这些线索从原来的无限多个可能中确定出一个词——含3个“y”的最短英文单词：syzygy。

超对称性也为满足这种对称性原理的理论提出了类似的约束。为认识这一点，假定我们现在遇到一个跟刚才那个语言问题类似的物理学难题。盒子里隐藏着某样东西——不知道是什么——具有一定的力荷。荷可能是电荷、磁荷或者别的什么更一般的荷。为具体起见，让

我们假定那是3个单位的电荷。如果没有进一步的信息，我们不可能确定盒子里的东西是怎么组成的。它可能是3个电荷为1的粒子，如3个正电子或3个质子；也可能是9个1/3电荷（如反下夸克）的粒子；还可能在这9个粒子之外另有任意数目的不带电荷的粒子（如光子）。就像只知道3个“y”的未知字母序列一样，盒子里有3个电荷的粒子组成也有无限多的可能。

这时候，像那字谜的情形一样，我们又听到两条线索：描述世界——包括盒子里的东西——的理论是超对称的，盒子里的东西是具有前面说的3个单位电荷的最小质量系统。通过波戈莫尼（E.Bogomol'nyi）、普拉萨德（Manoj Prasad）和索末菲（Charles Sommerfield）的发现，物理学家已经证明，具体明确的组织结构（这里如超对称的理论框架，在字谜的例子即英语的体系）和“极小性约束”（具有一定电荷的最小质量，或一定字母的最短单词）就意味着惟一确定了隐藏事物的本性。就是说，如果保证盒子里的东西是质量最轻的，并且还有确定的电荷，则物理学家就能完全确定它是什么东西。具有一定力荷的最小质量组成叫做BPS状态，是为了纪念它的3个发现者起的名字。[7]

BPS态的重要在于它的性质可以不借助微扰计算而简单、精确、惟一地确定。不论耦合常数是多少，它都是对的。就是说，即使弦耦合常数很大，微扰法不适用时，我们仍然可以导出BPS组成态的准确性质。这些性质通常叫非微扰的质量和力荷，因为它们的大小超越了微扰近似的框架。因为这一点，我们也可以认为BPS代表着“超越了微扰的状态”。

BPS性质只是大耦合常数下关于一定弦理论的整个物理学的一小部分，但它却让我们实在把握了某些强耦合的特征。当一个弦理论的耦合常数超过微扰论的适用范围时，我们就将有限的认识寄希望于BPS态。它们像几个恰当的外语单词，能把我们带得更远。

弦理论的对偶性

像惠藤那样，我们从一个弦理论说起，如I型弦；我们还假定9个空间维都是平直而非卷曲的。这当然不太现实，但可以使讨论简单一些，然后我们再说卷曲维的情形。我们从弦耦合常数远远小于1谈起。这种情况下，微扰论工具是行之有效的，它可以而且确实准确地算出了很多具体的理论性质。如果让耦合常数增大，但还是小于1，微扰方法仍然适用。不过，理论的具体性质会多少有些改变——例如，与两根弦的散射相关的数值结果可能不同，因为耦合常数增大时，图12.6的多圈过程会产生更大的影响。但除了具体数值的变化外，理论的物理内容还是一样的，只要耦合常数还在微扰论的界限内。

当I型弦理论的耦合常数超过1时，微扰法不能用了，我们只能去关心有限的非微扰质量和力荷的集合——BPS态——只有这一点还是我们能够认识的。惠藤讲的、后来经加利福尼亚大学波尔琴斯基（Joe Polchinsky）的合作研究证明的结果是：型弦理论的强耦合特征与杂化O型弦理论在小耦合常数下已知的特征是完全一致的。就是说，当I型理论的耦合常数很大时，我们能得到的质量和力荷特征正好等于从杂化O理论在小耦合常数下得到的那些特征。这强烈地暗示我们，看起来像冰与水那样全然不同的这两个弦理论，其实是对偶的。它提醒我们，I型理论在大耦合常数下的物理与杂化O理论在小耦合常数下的物理是完全相同的。相关的论证表明反过来也可能是对的：I型理论在小耦合常数下的物理与杂化O理论在大耦合常数下的物理也是完全相同的。[8]尽管两个理论在用微扰论方法分析时显得毫不相干，但现在我们看到它们（在耦合常数改变时）相互转变了——像冰与水的转变那样。

一个理论的强耦合物理可以用另一个理论的弱耦合图景来描绘，这个重要的新结果叫强弱对偶性。跟我们以前讲过的其他对偶性一样，它告诉我们那两个理论并不是迥然不同的。实际上，它们是同一

基本理论的不同描述。与中-英文的那个平凡对偶的例子不同，强弱对偶性是大有威力的。当两个对偶的理论中某一个的耦合常数小时，我们可以用充分发达的微扰方法来分析它的物理性质。如果理论的耦合常数很大，微扰方法不能用，我们现在也知道可以用对偶的图景来描述它——这里相关的耦合常数是小的，我们又可以用微扰论的工具了。这样的转换使我们能用定量的方法来分析原来认为超越了我们能力的理论。

不过，确实证明I型弦理论的强耦合物理等同于杂化O理论的弱耦合物理，是件极端困难的事情，现在还没有结果。原因很简单：对偶理论中的一方不能用微扰方法来分析，因为它的耦合常数太大了。这样，它的许多物理性质都不能直接计算出来。实际上，正是因为这一点，对偶性才更有潜力。因为，如果真是那样，则它为强耦合理论提供了新的分析工具：用微扰法去分析那个弱耦合的对偶图景。

但是，即使不能证明两个理论是对偶的，我们能满怀信心地确定的那些性质间的完美对应，却提供了令人不得不信的证据，说明我们猜想的I型与杂化O型弦理论间的强弱对偶关系是正确的。实际上，为检验这种对偶性，越来越精巧的计算都得到了肯定的结论。多数弦理论家相信，对偶性是真的。

用同样的方法，我们可以研究其余几个弦理论的强耦合性质，例如，IIB型弦理论。胡尔和汤森原来提出一个猜想，后来得到许多物理学家研究的支持，奇怪的事情果然发生了。当IIB型弦的耦合常数越来越大时，我们能认识的那些物理性质似乎跟IIB型弦本身的弱耦合情形完全相同。换句话说，IIB型弦是自对偶的。[9]具体地讲，详细分析揭示一个诱人的事实：当IIB型弦的耦合常数大于1时，如果我们将数值变换为它的倒数（这个值自然小于1），那么结果跟原来是完全一样的。跟我们在探索普朗克尺度下的卷缩维时发现的情形类似，如果把IIB型弦的耦合常数增加到大于1，自对偶性将证明那结果与原来耦合常数小于1的IIB型弦是完全等价的。

## 小结

现在来看我们都讨论了些什么。20世纪80年代中期，物理学家构造了5个不同的弦理论。在微扰论的近似框架下，这些理论显得各不相同。但近似方法只有在弦理论的耦合常数小于1时才适用。物理学家曾希望能计算每一个弦理论的耦合常数的精确数值，但那时能用的近似方程的形式不可能做到这一点。因此，物理学家便去研究每个理论在所有可能耦合常数值下的情形，小于1和大于1的情形——即弱耦合与强耦合。但传统的微扰方法对任何一个理论的强耦合特征都是无能为力的。

最近，物理学家借助超对称性的力量学会了如何计算一个弦理论的某些强耦合性质。令大多数圈内人士惊讶的是，杂化O型弦的强耦合性质似乎与I型弦的弱耦合性质是完全相同的，反过来也是。而且，IIB型弦的强耦合物理与它自身在弱耦合的情形相同。这些意外的关联激发我们沿着惠藤的路线走下去，看另外两个弦理论，IIA型与杂化E理论，是不是也能满足这样的图景。我们将遇到更加惊奇的事情。为做好准备，我们需要先简单回顾一下历史。

## 超引力

20世纪70年代末和80年代初，人们对弦理论还没有多大兴趣，许多理论物理学家还在点粒子量子场论的框架下寻求量子力学、引力和其他力的统一理论。他们看到了一点希望，那就是具有大量对称性的理论有可能克服点粒子的引力理论与量子力学间的矛盾。1976年，同在石溪纽约州立大学的弗里德曼（Daniel Freedman）、费拉拉（Sergio Ferrara）和纽文惠曾（Peter Van Nieu-wenhuizen）发现最有希望的是包含着超对称性的那些理论，因为玻色子和费米子消减量子涨落的趋势有助于平息微观世界的疯狂。他们用超引力来指那些想包容广义相对论的超对称量子场论。融合广义相对论与量子力学的这些努力最终都失败了。不过，如我们在第8章讲过的，物理学家从这

些探索中学会了一点教训，它们孕育着后来弦理论的发展。

那点教训经过法国巴黎高等师范学院克里默（Eugene Cremmer）、朱利亚（Bernard Julia）和谢尔克1978年的研究，变得再清楚不过了，那就是，最可能接近成功的是那些建立在更高维（而不是4维）空间的超引力理论。特别地说，最有希望的是10维或11维的形式。后来发现，11维的形式是最可能的。[10]与4个观测维的联系仍然在卡鲁扎和克莱茵的框架下实现：多余的维是卷缩的。在10维理论中，跟弦理论的情形一样，6维是卷缩的；而在11维理论中，7维是卷缩的。

当弦理论带着物理学家经过1984年的风暴时，点粒子超引力论的观点发生了巨变。我们曾反复强调过，当我们以今天或不远将来可能的精度来观察弦时，它看起来像一个点粒子。这种不太正规的说法还可以说得更准确一些：在研究弦理论的低能过程时——这些过程没有足够高的能量去探测超微观的弦的延展特性——我们可以运用点粒子量子场论的框架，将弦近似看成没有结构的点粒子。在面临短距离或高能量的过程时，我们不能再用这样的近似，因为弦的延展性是它能解决广义相对论与量子力学矛盾的关键，而点粒子理论做不到。不过，在足够低能的情形——距离足够大——不会遇到那些问题，我们为了计算的方便还常常用这种近似。

以这种方式最接近弦理论的量子场论不是别的，就是那个10维的超引力论。现在我们把20世纪七八十年代发现的10维超引力的特殊性质理解为弦理论基础力量下的低能“遗迹”。10维超引力的研究者们发现了冰山的一角，那角下藏着丰富的超弦结构。实际上，后来发现有4个不同的10维超引力理论，区别在于超对称性在理论中的具体作用方式。其中3个理论分别被证明是IIA、IIB和杂化E型弦的低能点粒子近似。另一个则同时表现为I型和杂化O型弦的低能点粒子近似；现在看来，那是这两个弦理论密切相关的第一条线索。

一切似乎都井然有序，但我们别忘了还有11维的超引力。建立在10维的弦理论显然没有空间容纳一个11维的理论。多年来，大多数（而不是全部）弦理论家抱有一种普遍的观点：11维的超引力不过是一个数学怪物，与弦理论的物理没有任何联系。[11]

**M理论是什么**

现在的观点不同了。在“95弦”年会上，惠藤论证说，如果从IIA型弦出发，把它的耦合常数从远小于1增大到远大于1，那么我们所能分析的物理（主要是饱和BPS态的组合）有一个低能的近似——那正是一个11维的超引力。

惠藤宣布这个发现时，在场的听众都惊呆了，从此也震撼着所有做弦理论的人。几乎弦领域的每一个人都感觉这是一个意想不到的进步。你对这个结果有什么第一反应呢？大概跟多数专家是一样的吧：一个确定的11维的理论怎么会与一个不同的10维理论相关呢？

答案有着深刻的意义。为理解这一点，我们先更准确地谈谈惠藤的结果。而实际上，更简单的办法是先说说惠藤和普林斯顿大学的一个博士后霍拉瓦（Petr Horava）后来发现的一个密切相关的结果，那是关于杂化E弦的。他们发现，强耦合的杂化E弦也有一个11维的图景，图12.7说明了那是为什么。在最左边的图，我们令杂化E弦的耦合常数远小于1。这是我们以前讨论过的情形，而弦理论家也研究过10多年了。从左向右，我们逐渐增大耦合常数，在1995年以前，弦理论家知道这样的结果是多圈过程（见图12.6）变得越来越重要；而随着耦合常数的增加，整个微扰论框架将最终失败。谁也不曾想过，当耦合常数增大时，一个新的维度也显露出来了！这是图12.7里的一个“垂直的”维度。别忘了，在这张图里，2维网格代表的是杂化E弦的整个9维空间。这样，垂直的新维是第10个空间维，它们与时间一起，构成一个11维的时空。

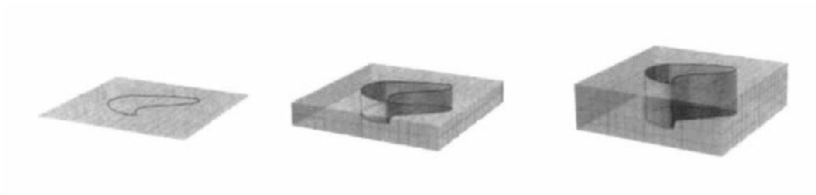


图12.7　随着杂化E弦耦合常数的增大，一个新的空间维出现了，弦本身也随之伸展成为柱形膜。

另外，图12.7还说明新维带来的一个深远结果。随着那一维的生长，杂化E弦的结构也在改变。当耦合常数增大时，它从1维的线圈伸展成一根丝带，然后成为一个变形的圆柱！换句话讲，杂化E弦实际上是一张2维膜，它的宽度（图12.7的垂向伸展）由耦合常数的大小决定。10多年来，弦理论家总是在用微扰论的方法，是一种建立在耦合常数很小的假设基础上的方法。正如惠藤所说，这样的假设使那些物质的基元表现得像一根根1维的弦。而实际上它们还有隐藏着的另一个空间维。从耦合常数很小的假设中解放出来，考虑杂化E弦在大耦合常数时的物理，那第2维就显露出来了。

这一发现并没有否定我们在前面几章下过的结论，但它迫使我们在新的框架下去认识它们。例如，这一切跟弦理论要求的1维时间和9维空间的图景如何相容呢？回想一下，从第8章我们知道，9维空间的约束条件来自弦能在多少个方向自由振动的问题，我们要求振动的方向数能保证量子力学几率有合理的数值。我们刚才发现的新维不是杂化E弦的振动方向，因为它是锁在“弦”本身的结构里的。换句话说，导出10维时空约束的微扰论方法从一开始就假定了杂化E弦的耦合常数很小。很久以后，人们才认识到，这必然得到两个相容的近似：图12.7的膜宽很小，从而看起来像一根弦；或者，第11维本来很小，超出了微扰方程的分辨能力。在这样的近似框架下，我们自然在头脑里形成一个充满着一维弦的10维宇宙。现在我们看到，那不过是包含着2维膜的11维宇宙的近似。

由于技术的原因，惠藤最先是在研究IIA型弦的强耦合性质时遇到

第11维的，情形与我们讲的类似。像杂化E弦的例子一样，这里第11维的大小由IIA型耦合常数决定。随着常数的增大，新的维也增大。不过，惠藤指出，在维增长中，IIA型弦不像杂化E弦那样伸展为丝带，而是形成图12.8那样的“内胎”。同样，惠藤又说，虽然理论家们总把IIA型弦看成只有长度没有粗细的1维物体，这只是假定弦耦合常数很小的微扰近似的反映。如果大自然真需要小的耦合常数，那么这种近似是值得信赖的。不过，惠藤和其他一些物理学家在第二次超弦革命中的研究强有力地表明，IIA型和杂化E的“弦”根本上说是存在于11维宇宙的2维膜。

那么，11维的理论是什么呢？在低能（与普朗克能量比）条件下，惠藤等人指出，人们忽略已久的11维超引力量子场论就是它的近似。但在高能条件下，我们又该如何描绘这个理论呢？这个问题如今还在积极研究中。我们从图12.7和图12.8知道，11维理论包含着2维延展的物体——2维膜。我们马上要讲，其他维的延展物也一样可能有重要作用。不过，除了不同性质的大杂烩以外，没人知道11维理论是什么。膜是它的基本物质组成吗？它的决定性特征是什么？它如何能够与我们了解的那些物理发生联系？如果相关的耦合常数很小，这些问题目前最好的答案就是我们在前面章节讲的那些，因为在小耦合常数时我们又回到弦理论。但如果耦合常数大，目前还没人知道结果会怎样。

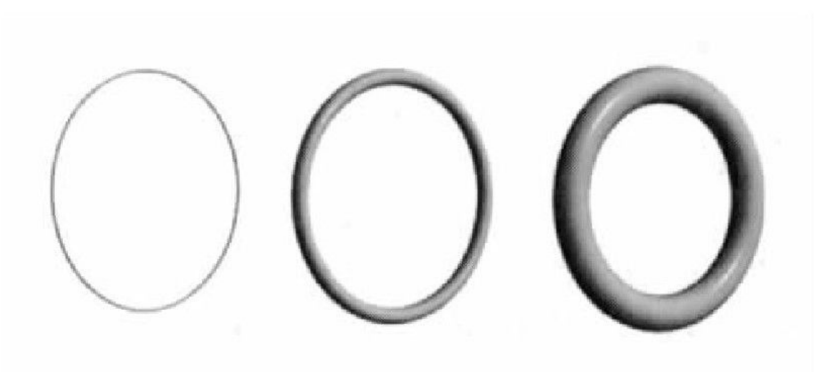


图12.8　耦合常数增大时，IIA型弦从1维线延展成为自行车内胎似的2维环状物体。

不管11维理论是什么，惠藤都暂时把它叫M理论。这名字代表很多意思，看你喜欢哪一个：谜一般的（Mystery）理论、母（Mother）理论（“一切理论之母”的意思）、膜（Membrane）理论（因为不论结果如何，膜似乎都是理论的一部分）、矩阵（Matrix）理论（这个名字是根据鲁杰斯（Ratgers）大学邦克斯（Tom Banks）、奥斯汀得克萨斯大学的费施勒（Willy　Fischler）、鲁杰斯大学的申克（Stephen　Shenker）和苏斯金为理论提出的一种新解释而提出的）。但是，即使不了解它的名字，没严格把握它的性质，我们还是清楚地知道，M理论为把5个弦理论结合在一起提供了统一的基础。

M理论与对偶网

有一个古老的寓言，讲的是3个盲人和1头大象的故事。第一个盲人抓住了象牙，就说它又尖又滑；第二个盲人抱住一条腿，说它是粗壮结实的柱子；第三个盲人拖着尾巴，说它是纤细有力的鞭子。3个人说的截然不同，而谁也看不见别人，所以都以为自己抓住的是不同的动物。多年来，物理学家也像盲人那样在黑暗里摸索，认为那些不同的弦理论本来就是不同的。但现在经过第二次超弦革命的发现，物理学家认识到M理论就是统一5个弦理论的那头大象。

我们在这一章已经讨论过由于超越微扰论框架——本章之前实际上一直在微扰论的框架下——而带来的对弦理论认识的改变。图12.9总结了我们到目前为止所发现的一些关系，箭头指对偶理论。你可以看到，我们有一个关联网，但还不完整。把第10章的对偶性也包括进来，我们就能把它完成。




图12.9　箭头说明哪两个理论是对偶的。

回想一下大-小半径的对偶性（以半径1/R替代R）。以前我们忽略了这种对偶性的一个方面，现在我们来说明它。在第10章，我们讨论弦在一个具有圆周维的宇宙中的性质，但没有具体说明我们用的是5个弦理论中的哪一个。我们说，变换弦的缠绕和振动模式后，我们可以用圆周维半径为R的宇宙的弦理论来同样准确地描述半径为1/R的那一个。我们忽略的一点是，IIA和IIB型理论在这个对偶性下实际发生了转换，杂化O和杂化E弦也是这样。就是说，大-小半径对偶性的更准确表述应该是在圆周维半径为R的宇宙中的IIA型弦的物理完全等同于圆周维半径为1/R的宇宙中的IIB型弦的物理（类似的表述对杂化E和O弦也是成立的）。对大-小对偶性的这种修正，并不影响第10章的结论，但对我们现在的讨论却有着重要影响。

原来，当IIA和IIB型弦理论以及杂化E和杂化O理论间的联系建立起来后，大-小半径的对偶性便完成了我们说的联系网，如图12.10的虚线。这图说明所有那5个弦理论连同M理论都是相互对偶的。它们都嵌入了一个理论框架；它们提供了描述同一基本物理的5种不同的途径。在某些情形，一种表述可能比另一种表述有效得多。例如，处理弱耦合的杂化O理论就比处理强耦合的I型弦容易得多。不过，它们描写的完全是同一种物理。

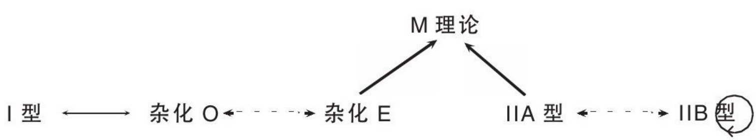



图12.10　把时空几何形式（第10章）的对偶性包括进来，所有5个弦理论和M理论就在一个

对偶网中联结在一起了。

宏图

现在，我们可以更完整地来认识图12.1和图12.2了——那是我们在这一章的开头为了概括基本要点而引进的两个图。在图12.1中我们看到，1995年以前，在没有任何对偶性考虑时，我们有5个显然不同的弦理论。不同的物理学家抱着一个理论，由于不知道对偶性，这些理论看起来是不同的。每一个理论都有变化的性质，如耦合常数的大小，卷缩维的几何形式和大小。物理学家曾经（现在也仍然）希望能从理论本身来确定这些决定性的性质，但现在的近似方程却没有能力做到这一点，所以他们自然去研究各种可能出现的物理。这是图12.1中以阴影表示的区域——区域内每一点表示一种特别的耦合常数和卷缩维几何的选择。没有对偶性，我们仍然只有5个脱节的理论（集合）。

但是现在，如果把前面讨论过的所有对偶性都应用进来，另外还包括那个统一的M理论的中心区域，那么我们就能随着耦合常数和几何参数的改变，从一个理论转换到另一个理论；这就是图12.2所表示的内容。即使我们对M理论没有多少认识，这些间接的论证也令我们强烈感到，它为5个原来显得不同的弦理论提供了统一的基础。而且，我们也知道，M理论还紧密联系着另一个理论——11维超引力论——这画在图12.11，它比图12.2更准确一些。[12]

图12.11说明M理论的基本思想和方程（尽管目前只有部分了解）统一了所有的弦理论思想和方程。M理论像一头理论的大象，令弦理论家们睁开了双眼，看到了一个更宏大的统一框架。

M理论的奇异特征：膜的民主

当弦耦合常数很小时，图12.11中上面5个伸出的触角区域的弦理论的基本物质组成都表现为1维的弦。然而，我们刚得到一个新发

现。如果从杂化E或IIA型区域出发，增大各自的耦合常数值，我们将走进图12.11的中心区域，原来1维的弦将展开成2维的膜。而且，经过一系列对偶关系的转换——包括弦耦合常数和卷缩空间维的具体形式——我们能自由连续地从图12.11的一点转移到另一点。从杂化E和IIA型弦生成的2维膜，也可以在我们向其他3个弦理论的转移中生成，于是我们看到，5个弦理论都包含着2维的膜。

这引出两个问题。第一，2维膜是弦理论真正的基本组成吗？第二，我们在20世纪70年代和80年代初从零维点粒子跳跃到1维弦，现在又看到弦实际上是2维膜，那么在理论中还会有更高维的物质组成吗？我写这些问题时，还没有完全的答案，不过可能是下面的情形。

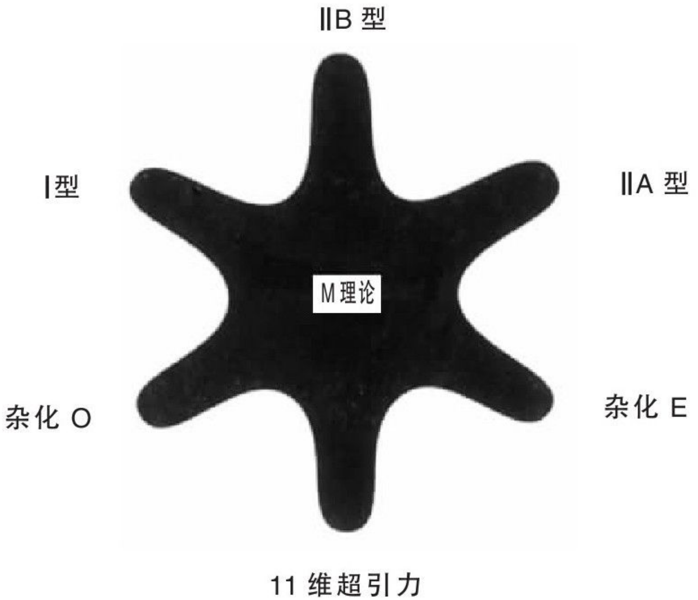



图12.11　把对偶性包括进来，5个弦理论和11维的超引力以及M理论就在一个统一框架下结合在一起了。

在微扰论近似成立的范围外，我们主要依靠超对称性来认识每个弦理论的某些性质。特别是BPS态的性质，它们的质量和力荷，是由

超对称性惟一决定的，这使我们不经过艰难的直接计算就能认识它们的某些强耦合特征。实际上，经过霍罗维茨和斯特罗明戈的原始研究和后来波尔琴斯基的奠基性工作，我们现在对BPS态懂得更多了。特别是，我们不仅知道它们携带的质量和力荷，还清楚地知道它们像什么。它们的图像也许是所有发现中最令人惊奇的。有些BPS态是1维的弦，有些是2维的膜，这都是我们所熟悉的。令人惊奇的是还有3维、4维的——实际上，任何空间维都是可能的，包括9维。弦理论或M理论或别的什么最后的理论，实际上包含着具有任何可能空间维数的延展物。物理学家用3-膜来称具有3个空间维的物体，4-膜则具有4个空间维，一直到9-膜（更一般地说，对一个具有p个空间维的物体（这里p是一个整数），物理学家找了一个更有韵味的名字：p-膜）。用这些名词，有时我们说弦是1-膜，寻常的膜为2-膜。所有这些延展事物都是理论的一部分，于是，汤森说这是“膜的民主”。

不论有多少平等的“膜”，弦这1维的延展物却是与众不同的。原因是这样的：物理学家已经证明，除了1维的弦而外，不论在图12.11的哪个弦理论中，不同维的物体的质量都反比于相关耦合常数的值。这意味着，在弱耦合时，任何一个理论中除弦以外的所有事物都是大质量的——数量级大于普朗克质量。因为质量大，从而$E=mc^2$的能量也大，所以膜对许多（但不是所有，我们很快要在下一章讨论）物理的影响是很微弱的。但是，当我们大胆走出图12.11的触角区域时，高维的膜将变轻，而它的影响将变大。[13]

于是，我们应该牢记这样一幅图景：在图12.11的中央区域，理论的基本物质组成不仅有1维的弦、有2维的膜，还有不同维数的高维“膜”，它们几乎都是平等的。目前，这个完全理论的许多基本特征我们还没有严格把握，但我们能肯定一件事情：当我们从中央转移到边缘任何一个触角区域时，只有1维的弦（或者像图12.7和图12.8中卷缩起来更像弦的膜）才足够轻，才能与我们熟悉的世界——如表1.1里的粒子和它们相互作用的4种力——发生联系。弦理论家们用了

近20年的微扰方法还没有能力揭示那些超大质量的高维延展物的存在；弦主宰着我们的分析，所以理论的名字还是离“民主”十分遥远的“弦理论”。在图12.11的边缘区域，我们又一次证明了，在大多数情况下，除了弦以外，别的都可以忽略。根本说来，本书到目前为止都是那么做的。不过，我们现在明白了，理论实际上比以前任何人想象的都丰富得多。

那些东西能回答弦理论未解决的问题吗

能，也不能。我们设法从某些结论摆脱出来——现在看来，那些结论不过是微扰近似分析的一些结果，而不是真正的弦理论的结果——从而深化了我们的认识。但我们今天的非微扰工具的能力还太有限。对偶关系网的发现让我们更深入地认识了弦理论，但还有很多问题没有解决。例如，我们现在还不知道如何超越弦耦合常数的近似方程——我们已经看到，那些方程太粗了，得不出什么有用的信息。我们也还不明白为什么正好有3个展开的空间维，也不知道该如何选择卷缩维的具体形式。这些问题需要比我们现有的磨得更加锋利的工具才能解决。

我们确实把握的，是更深入地认识了弦理论的逻辑结构和理论范围。在图12.11总结的认识之前，每个理论的强耦合行为还是一只黑箱，一个无人知晓的谜。强耦合的区域像老地图上的一块处女地，那里可能潜藏着巨龙和海怪。不过现在我们看到，尽管通向强耦合的旅程会带我们穿过陌生的M理论的领地，但它最终还是让我们舒适地躺在弱耦合的怀抱里——尽管在对偶的语言下，那也曾被认为是不同的弦理论。

对偶性和M理论统一了5个弦理论，它们还提出一个重要结论。我们未来的发现也很可能没有比刚才讲的那些更令人惊奇的了。如果哪位地图专家能填满地球表面的每一个角落，地图就画完了，地理学知识也到头了。这并不是说南极探险或密克罗尼西亚孤岛旅行没有科

学和文化的意义，而只是说地理大发现的时代结束了。全球没有一个空白点，当然也没有什么需要去“发现”的。对弦理论家来说，图12.11的“理论地图”扮演着类似的角色。从5个弦理论的任何一个开始扬帆远航，都走不出它所覆盖的理论区域。虽然我们还远未完全弄清M理论环球远行的路线，但地图上已经没有空白点了。弦理论家现在可以像地图专家那样满怀自信地宣布，过去百年的基本发现——狭义和广义相对论，量子力学，强、弱和电磁力的规范理论，超对称性，卡鲁扎和克莱茵的多维空间……——从逻辑上说，都完全包容在图12.11的理论中了。

弦理论家——也许应该说M理论家——面临的挑战，是证明图12.11的理论地图上的某个点确实描绘了我们的宇宙。这需要寻找完整而准确的方程，让它的解去捕捉图中那个飘忽不定的点，然后以足够的精度去理解相应的物理，从而与实验结果进行对比。正如惠藤讲的：“认识M理论究竟是什么——它赋予的物理是什么——至少会像历史上的任何一次伟大的科学变革一样，极大地改变我们对自然的认识。”[70]这是21世纪物理学大统一的纲领。

注释

1.我们简单概括一下5个弦理论之间的差别。为此，我们注意沿弦圈的振动扰动可以是顺时针的，也可以是反时针的。IIA和IIB型弦的差别在于，在IIB型理论中，顺、反时针的振动是一样的，而在IIA型理论中，两个方向的振动正好相反。在这里，“相反”有着准确的数学意义，不过可以简单地用每个理论的弦振动模式的自旋来理解。在IIB型理论中，所有粒子在同一方向上自旋（它们具有相同的手征性），而在IIA型理论中，粒子在两个方向上自旋（具有两种手征性）。尽管如此，两个理论都包含着超对称性。两个杂化理论的差别也基本是这样，但差别更大。它们的顺时针弦振动看起来跟两个II型弦的情形一样（只考虑顺时针振动时，IIA和IIB型理论是相同的），但它们的反时针振动却是原始的玻色弦理论的情形。尽管同时考虑玻

色弦的顺时针和反时针振动会遇到难以逾越的障碍，但在1985年，格罗斯（David Gross）、哈维（Jeffrey Harvey）、马丁尼克（Emil Martinec）和罗姆（Ryan Rhom）（四个人那时都在普林斯顿大学，绰号叫“普林斯顿弦乐四重奏”）证明，如果把它跟II型弦结合起来，则我们能得到一个非常合理的理论。这种结合真正奇怪的地方是，玻色弦需要26维时空——这是一个老结果，自鲁特杰斯大学劳弗莱思（Claude Lovelace）1971年的研究以及波士顿大学布罗维尔（Richard Brower）、剑桥大学戈达（Peter Goddard）和盖恩斯维尔弗罗里达大学索恩（Charles Thorn）1972年的工作，我们就知道它了——而超弦如我们讲的只需要10维时空。所以，杂化弦理论的结构是一种奇特的“杂交”的东西——一种具有“杂交优势（heterosis）”的产物——反时针振动的弦在26维里活动，而顺时针振动的弦却活动在10维！你大概还没太明白这令人困惑的杂交是怎么回事。格罗斯和他的伙伴们已经证明，玻色弦那多出的16维一定卷缩成一个或两个特别高维的面包圈的样子，从而生成杂化O和杂化E理论。由于玻色弦那多余的16维是紧紧卷缩在一起的，所以这两个理论就像II型理论那样，表现为只有10维的样子。当然，两个杂化的理论还是具有超对称性的某种形式。最后，I型理论是IIB型理论的“亲戚”，不过，它除了有我们在前面章节里讨论过的闭弦外，还有两端没有联结的所谓开弦。

2.我们这一章里说的“精确”答案，如地球的“精确”运动，实际上指的是在某个选定的理论框架内对某些物理量做出准确的预言。在我们真正拥有一个“最后的”理论以前——也许我们今天就有了，也许永远也不可能有——我们的一切理论就自身说来都是实在性的某种近似。但近似的概念与我们本章讨论的东西无关。我们这里关心的是这样一个事实：在一个选定的理论中，常常很难（虽然不是不可能）得到精确的理论预言。我们只得用以微扰理论为基础的近似方法来得出那些预言。

3.这些图是所谓费曼图的弦理论形式。费曼图是费曼为在点粒子

量子场论中进行微扰计算而发明的。

4.更准确地说，每一虚弦对（即给定图中的每一个圈）都为弦耦合常数增加了一个自乘因子（当然它们还有别的更复杂的贡献）。圈越多，弦耦合常数的自乘因子越多。如果弦耦合常数小于1，多次自乘将使总的贡献更小；如果常数等于或大于1，则多次自乘的结果是1或者远远大于1。

5.对数学感兴趣的人可以看到，这个方程说明时空允许平直的Ricci度规。如果把时空分解为4维闵可夫斯基时空和6维紧致Kähler空间的笛卡儿积，则Ricci平直性等价于Kähler空间是一个卡-丘流形。这就是为什么卡-丘空间在弦理论中起着那么举足轻重的作用。

6.当然，没有什么绝对保证这些直接方法是可靠的。例如，在宇宙其他遥远的区域，物理学定律也可能不同，就像人的脸也有左右不对称的。我们将在第14章简单讨论这个问题。

7.专业的读者会发现这些结果要求所谓N＝2超对称性。

8.说得更准确一点，假如，杂化O耦合常数为g HO，I型耦合常数为gI，则两个理论间的关系说的是，只要g HO＝1/gI或者gI＝1/gHO，那么它们在物理上就是相同的。一个耦合常数大时，另一个耦合常数就小。

9.这很像前面讨论的R-1/R对偶性。如果IIB型弦耦合常数为gIIB，则gIIB和1/gIIB很可能描写同一样的物理。如果gIIB大，则1/gIIB小，反之亦然。

10.假如只有4维是卷缩的，则总维数超过11的理论必然生成自旋大于2的无质量粒子，这是实验和理论都排斥的东西。

11.一个值得注意的例外是杜弗、霍维（Paul Howe）、稻见赳夫（Takeo I-nami）和斯特勒（Kelley Stelle）在1987年的一项重要工

作。他们借助贝格雪夫（Eric Bergshoeff）、塞金（Ergin Sezgin）和汤森的发现，证明11维弦理论应该具有深层的11维联络。

12.更准确地讲，这个图应该解说成我们有一个依赖于大量参数的理论。参数包括耦合常数、几何大小和形态参数。原则上说，我们应该能用理论来计算每一个参数的数值——如耦合常数的值和一定时空几何的具体形式——但凭我们目前的理论认识，还不知道如何实现这一点。所以，为了更好认识这个理论，弦理论家研究当参数在一切可能范围内变化时，理论表现出什么性质。假如选择的参数值落在图12.11那6个边缘区域内，那么理论将表现出5个弦理论或11维超引力论所固有的性质，这一点我们讲过了。如果选择的参数值在中央区域，那么物理就是还像谜一样的M理论所统治的物理。

13.然而我们应该知道，即使在边缘区域，高维的膜也可能以某些奇异方式影响我们寻常的物理。例如，有人提出，我们的3个空间维可能本来就是一个巨大展开的3-膜。如果真是那样，我们每天的生活就都是在一个3维膜的内部度过的。现在正有人在考察会不会有这样的事情。

## 第13章　从弦/M理论看黑洞

弦理论出现以前，广义相对论与量子力学间的矛盾真把我们的直觉大大地羞辱了一回——我们一贯直觉地认为，自然律应该是天衣无缝的一个和谐的整体。而那矛盾还不仅仅是理论上的一道巨大裂缝。如果没有引力的量子力学体系，我们不可能认识发生在宇宙大爆炸时刻和统治着黑洞内部的那些极端的物理条件。随着弦理论的发现，我们今天有希望揭开这些深藏的秘密了。在这一章和下一章里，我们要讲弦理论朝着认识黑洞和宇宙起源的方向走了多远。

### 黑洞和基本粒子

乍看起来，很难想象还有哪两样东西能比黑洞和基本粒子有更大

的差别。我们常把黑洞描绘成天体的巨无霸，而基本粒子却是物质的小不点儿。但20世纪60年代末和70年代初的许多物理学家，包括克里斯托多罗（Demetrios Christodoulou）、伊思雷尔（Werner Israel）、普赖斯（Richard Price）、卡特尔（Brandon Car-ter）、克尔（Roy Kerr）、罗宾森（David Robinson）、霍金和彭罗斯，发现黑洞和基本粒子也许不像我们想的那么悬殊。他们发现越来越多的证据令人相信惠勒所谓的“黑洞无毛”所表达的思想。惠勒这话的意思是，除了少数可以区别的特征外，所有黑洞看起来都是相像的。那几个可以区别的特征，第一当然是黑洞的质量。别的呢？研究发现它们是黑洞所能携带的电荷或其他力荷，还有它的自转速度。就是这几样。任何两个黑洞，如果有相同的质量、力荷和自转，它们就是完全相同的。黑洞没有炫目的“发型”——就是说，没有别的内在的特征——将自己区别出来。这情形我们似曾相识——别忘了，正是这些性质、质量、力荷和自旋，将基本粒子彼此区别开来。因为在决定性特征上的相似，许多物理学家这些年来形成一个奇特的猜想：黑洞可能本来就是巨大的基本粒子。

实际上，根据爱因斯坦的理论，黑洞没有极小质量的限制。任何质量的一团物质，如果被挤压得足够小，我们能直接用广义相对论证明它可以成为一个黑洞。（质量越小，我们就把它压得越小。）这样，我们可以想象一个思想实验：从质量越来越小的小块物质开始，我们把它们压成越来越小的黑洞，然后拿这些黑洞与基本粒子进行比较。惠勒的“无毛”结论令我们相信，如果质量足够小，我们以这种方式形成的黑洞看起来很像基本粒子。两样小东西都完全由它们的质量、力荷和自旋来刻画。

但有一个问题。天体物理学的黑洞，质量是太阳的许多倍，既大且重，量子力学与它们没有关系，只需要用广义相对论来理解它们的性质。（这里讲的是黑洞的整个结构，没考虑黑洞中心的坍缩奇点，那个小东西当然是需要量子力学来描述的。）然而，当我们形成越来

越小的黑洞时，可能出现量子力学确实发生作用的情形。例如，当黑洞总质量为普朗克质量或更小的时候。（从基本粒子物理学的观点看，普朗克质量是巨大的——约质子质量的1000亿亿倍。但从黑洞的观点看，普朗克质量是相当小的，不过等于一粒灰尘的质量。）于是，猜想小黑洞与基本粒子密切相关的物理学家迎面就碰上广义相对论这一黑洞的理论核心与量子力学的不相容问题。过去，两者的不相容曾死死地拖着人们向前的脚步。

弦理论能让我们往前走吗

是的。通过黑洞的一个喜出望外的大发现，弦理论在黑洞与基本粒子间建起了第一个合理的理论联系。通往联系的道路是曲折的，但它会领着我们经过弦理论的一些最有趣的发展，是一段令人难忘的历程。

事情从20世纪80年代末以来弦理论家一直谈论的一个看似毫不相干的问题开始。物理学家和数学家很早就知道，当6个空间维卷缩成卡-丘形式时，在空间结构中一般存在两种类型的球面。一种是2维的，像沙滩皮球的表面，在第11章的空间破裂翻转变换中起着积极的作用；另一种很难想象，但同样是普遍存在的，那就是3维球面——在有4个展开的空间维的宇宙中，海滩上玩的就该是这样的皮球。当然，正如我们在第11章讲的，我们世界的普通的沙滩皮球本来也是3维的东西，但它的表面，就像花园里浇水管子的表面一样，是2维的。我们只需要两个数——如经度和纬度——就能确定表面上任何一点的位置。但我们现在是在想象多一个空间维的情形：一个4维的沙滩皮球，它的表面是3维的。这样的皮球我们几乎不可能在头脑里画出来，所以在大多数时候我们还是会借助更容易“看得见”的低维类比来想象它。不过，我们马上会看到，这多1维的球面有一点性质是至关重要的。

通过对弦理论方程的研究，物理学家发现，随着时间的演化，3

维的球面可能而且极有可能收缩——坍缩——下去，直到几乎没有体积。那么，弦理论家问，如果空间结构这样坍缩下去，会发生什么事情呢？空间的这种破裂会带来什么灾难性的后果吗？这很像我们在第11章提出并解决了的问题，但那里我们只考虑了2维球面，而现在我们面对的是3维球面的坍缩。（在第11章，我们想象卡-丘空间的一部分收缩，而不是整个空间都收缩，所以第10章的大小半径的等同性不适用了。）维数的不同带来了性质上的根本差异。[1]回想一下我们在第11章讲过的东西。当弦在空间移动时，它们能“套住”2维的球面。就是说，弦的2维世界叶能像图11.6那样把2维球面完全包裹起来。可以证明，这足以避免坍缩、破裂的2维球面可能产生的物理学灾难。但是，我们现在面临着卡-丘空间里的另一类球面，它的维太多，一根运动的弦不可能把它包围起来。如果你觉得这一点不好懂，请你考虑一个类似的低维的例子。你可以把3维球面想象成普通的2维沙滩皮球的表面，不过同时，你还得把1维的弦想象成0维的点粒子。这样，你可以看到，零维的点粒子什么也套不住，当然更套不住2维的球面；同样，1维的弦也不可能套住3维的球面。

这种思路引导弦理论家们猜想，假如卡-丘空间里的3维球面要坍缩——近似方程表明这是很可能（即使不是很普遍）在弦理论中发生的事情——那么它可能会带来灾难性的结果。实际上，90年代中期以前发展起来的近似弦理论方程似乎说明，假如那样的坍缩发生了，宇宙的活动可能会慢慢停歇下来；那些方程还意味着，某些被弦理论控制了的无限大将重新被那样的空间破裂“解放”出来。多年来，弦理论家们不得不生活在这样恼人的没有结果的思想状态下。但在1995年，斯特罗明戈证明，那些绝望的论调和猜想都是错误的。

跟着惠藤和塞伯以前的奠基性工作，斯特罗明戈发展了弦理论的新认识，那就是，以第二次超弦革命的新眼光来看，弦理论并不仅仅是1维的弦的理论。他的思路是这样的：1维的弦——用新的术语讲，即1-膜——能完全裹住一块1维的空间，如图13.1的一个圆圈。（注

意，这图跟图11.6不同，那里是1维的弦在运动中套住一个2维的球面。图13.1应看作是某一瞬间的镜头。）同样，我们在图13.1看到，2维的膜能卷起来完全包裹一个2维球面，就像一张塑料膜紧紧包裹一个橘子。虽然那很难想象，斯特罗明戈还是沿着这条思路发现，弦理论中新出现的3维物质基元——3-膜——能卷曲并完全覆盖3维的球面。看清这点后，他接着用简单标准的物理计算证明，卷曲的3-膜仿佛一个特制的盾牌，完全消除了弦理论家们害怕在3维球面坍缩时可能发生的灾难。

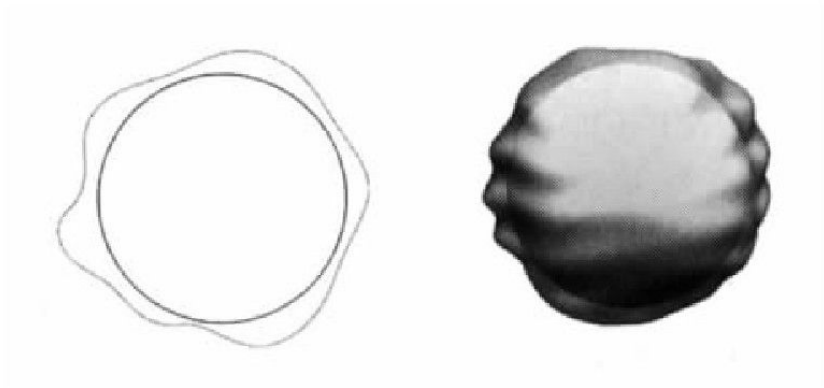


图13.1　一根弦可以包围卷起来的一片1维空间；2维膜可以卷起来包裹一块2维表面。

这是奇妙而重要的发现，但它的力量要过些时候才能完全显露出来。

撕裂空间结构

物理学最激动人心的事情是在一夜之间发生认识的改变。斯特罗明戈在互联网上发布他论文的第二天早晨，我就在康奈尔的办公室里从www上看到它了。他用弦理论的新的激动人心的发现一举解决了关于额外维卷缩成一个卡-丘空间的棘手难题，不过，在我思考他的文章时，觉得他可能只做了一半的事情。

在第11章讲空间破裂翻转变换的现象时，我们研究了两个过程：2维球面收缩成一个点，使空间发生破裂，然后球面又以新的方式膨胀，从而修复裂痕。在斯特罗明戈的文章里，他分析了3维球面收缩成一点的过程，证明弦理论新发现的高维物体将确保物理学过程继续良好地进行下去。到这里，他的文章就结束了。他是不是忘了，也许还有事情的另一半——破裂的空间通过球面的重新膨胀而修复？

1995年春，那时莫里森正在康奈尔访问我，那天下午我们一起讨论了斯特罗明戈的论文。两三个小时后，我们对“事情的另一半”有了一个轮廓。根据20世纪80年代末以来数学家们的一些研究成果——那些数学家包括，犹他大学克里门斯（Herb Cle-mens）、哥伦比亚大学弗里德曼（Robert Friedman）、沃威克大学雷德（Miles Reid）——以及坎德拉斯、格林和胡布施（Tristan Hübsch）（那时都在奥斯丁得克萨斯大学）的应用，我们发现，当3维球面坍缩时，卡-丘空间也许可能破裂然后通过球面膨胀而再复原。但是这里出现了很奇怪的事情：坍缩的球面是3维的，而新膨胀起来的球面只有2维。很难具体把它的样子画出来，不过我们可以从低维类比中得到一点认识。我们不去考虑那个令人难以想象的3维球面坍缩然后被一个2维球面取代的情形，让我们来想象一个1维球面的坍缩，然后它被一个0维球面所取代。

首先，什么是1维和0维的球面呢？让我们用类比来说明。2维球面是3维空间里的一个点集，集合中每一点到一个选定的中心的距离都相同，如图13.2（a）。根据同样的思想，1维球面是2维空间（如本页的表面）里的点集，每一点到某个中心有相同的距离。如图13.2（b）所示，其实它就是一个圆周。最后，根据这样的方式，0维球面是1维空间（直线）里到某中心等距离的点的集合。如图13.2（c）所示，0维的球面只有两个点，它的“半径”等于每点到公共中心的距离。这样，上面指的低维类比说的是一个圆（1维球面）收缩，然后破裂，接着成为两个点（0维球面）。图13.3实现了这个抽

象的过程。

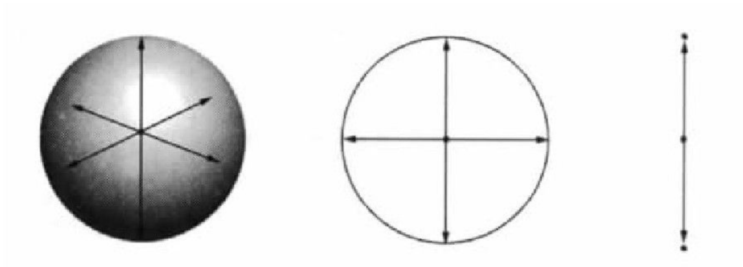



图13.2　看得见的几种球面——（a）2维，（b）1维，（c）0维。

我们从一个面包圈的表面开始，它当然包含着1维的球面（圆），图13.3突出了一个。现在我们想象，随时间流逝，图中那个圆开始坍缩，引起空间结构收缩。我们可以像下面那样来修复陷落的空间结构：让它在瞬间破裂，然后用0维的球面（两个点）取代原来收缩的1维球面（圆）来弥合破裂生成的上下两个洞。如图13.3，这样的结果像一只弯曲的香蕉，通过轻微的变形（没有空间破裂），它可以再形成一个光滑的沙滩皮球样的表面。于是我们看到，当1维球面坍缩并被0维球面取代时，原来面包圈的拓扑（即它的基本形状）会发生巨大改变。在卷缩的空间维的情形，图13.3的空间破裂过程将使图8.8的宇宙演化成为图8.7的宇宙。

尽管这是一个低维类比，但在我们看来，它还是抓住了莫里森和我为斯特罗明戈设想的“事情的另一半”的基本特征。卡-丘空间里的3维球面坍缩以后，空间会破裂，然后它生成一个2维球面来修复自身，那将导致剧烈的拓扑改变，比惠藤和我们在以前的研究（11章讨论的）中发现的那些变化可怕得多。这样，从根本上说，一个卡-丘空间可以将自己变换成另一个形态完全不同的卡-丘空间——就像图13.3的面包圈变成沙滩皮球一样——而弦物理学在变化中仍然保持着良好的表现。虽然显露了一点风光，但我们知道还有许多重要的方面需要考虑，把它们都弄清楚了，我们才能肯定我们的“事情的另一半”不会

带来任何奇怪的东西——令人厌恶的和物理上不能接受的结果。那天晚上，我们各自带着一时的欢喜回家了——欢喜我们有了一个新的重大发现。



图13.3　面包圈（环）上的一个圆周坍缩成一点，表面从那点分裂，生成两个破裂的洞。一个0维球面（两点）用来“黏合”它，从而取代了原来的1维球面（圆），修复了破裂的表面。这样的过程使原来的面包圈变成了形状完全不同的沙滩皮球。

**E-mail风**

第二天早晨，我收到斯特罗明戈的电子邮件，问我对他的文章有什么评论或反应。他说“它在某种程度上应该是与你同阿斯平沃尔和莫里森的工作有关的”，因为，后来我们知道，他也曾探索过它跟拓扑改变现象有什么可能的联系。我立即给他回了信，把莫里森和我刚得到的蓝图向他大概描绘了一下。从回答看，他的兴奋显然跟莫里森和我昨天的心情是一样的。

接下来的几天里，电子邮件如流水似的在我们3个人之间流淌，我们在狂热地寻求将空间破裂的拓扑改变思想严格定量地表达出来。所有的细节都逐渐确定下来了。到星期三，也就是斯特罗明戈发表他发现的一个星期后，我们联合论文的稿子已经写好，它揭示了随3维球面的坍缩而出现的一种新的巨大的空间结构变换。

斯特罗明戈原定在第二天去哈佛演讲，所以一早就离开圣巴巴拉了。我们说好由莫里森和我继续修饰论文，然后在当天晚上在电子档案上发表。夜里11：45，我们把计算反复校核后，觉得没有问题了，便把文章发出去；我们也走出物理系的大楼。莫里森和我向我的车走

去（他自己在访问期间租了房子，我想送他回去），我们的讨论也变得有点儿吹毛求疵。我们在想，如果有人不想接受我们的结果，最刺耳的批评会是什么样的？我们驱车离开停车场，驶出校园时，才发现尽管我们的论证很有说服力，但也不是完全无懈可击的。我们谁也没想过它可能会错，但确实感到在文章的某些地方，我们下结论的语气和特别的用词可能会招惹不愉快的争论，从而淡化了结果的重要性。我们都觉得文章本来可以做得更好一些：调子放低一点，结论下得温和一点，让物理学同行自己去判断文章的优劣，而不是让它像现在那个样子去惹人反感。

车往前开着，莫里森提醒我，根据电子档案的规则，我们可以在凌晨2：00以前修改我们的文章，过后它才在公共网上发表。我立即调转车头，开回物理楼，撤回原来的文章，开始降低它的语调。谢天谢地，这做起来很容易。在评论的段落里改换几个词，就把锋芒藏起来了，一点儿也不影响技术内容。不到1个钟头，我们又把它发出去，并且说好在去莫里森家的路上，谁也别再谈它。

第二天刚下午时，我们就发现文章引起的反响是很热情的。在众多回信里有一封来自普里泽，他把我们大大恭维了一番，那可能是一个物理学家对别人最大的公开的赞美："如果我能想到那一点就好了！"虽然头一天晚上还在担心，但我们让弦理论家们相信了，空间结构不仅能发生以前（第11章）发现的那种"温和的"分裂，像图13.3简单描绘的那种暴烈得多的破裂，也同样可能发生。

重回黑洞和基本粒子

上面讲的东西与黑洞和基本粒子有什么关系呢？关系可多了。为认识这一点，我们还得问自己一个我们曾在第11章提出过的问题：那样的空间破裂产生了什么可观测的物理结果？我们已经看到，对翻转变换来说，答案令人惊讶：什么也没发生。对于我们新发现的这种剧烈空间破裂变换——所谓的锥形变换（co-nifold transition）——结

果还是一样的，没有传统广义相对论会出现的物理学灾难，但有更显著的看得见的结果。

这些看得见的结果背后有两个相关的概念，我们一个一个来解释。首先，如我们讲的，斯特罗明戈的开拓性突破是他发现卡-丘空间里的3维球可能发生坍缩，却不会带来灾难，因为“裹”在它外面的3-膜提供了理想的保护层。但那卷曲的膜像什么样子呢？答案来自霍罗维茨和斯特罗明戈以前的研究，他们曾证明，对我们这些只认识3个展开的空间维的人来说，“涂”在3维球面上3-膜将产生一个引力场，看起来像一个黑洞。[2]这不是一眼能看出来的，只有详细研究了膜的方程以后才能弄清楚。可惜，这种高维图像仍然很难画在纸上，不过图13.4用2维球面的低维类比传达了大概的意思。我们看到，2维的膜可以涂抹在2维球面上（球面本身也处在展开维的某个地方的卡-丘空间里）。如果有人通过展开维向那个地方看去，他会通过卷曲膜的质量和力荷而感到它的存在。据霍罗维茨和斯特罗明戈的证明，那些性质就像是一个黑洞的性质。而且，斯特罗明戈在他1995年的突破性论文里还证明了3-膜的质量——也就是黑洞的质量——正比于它所包围的3维球面的容积：球面容积越大，包裹它的3-膜就越大，从而质量也越大。同样，球面容积越小，包裹它的3-膜的质量也越小。于是，当球面收缩时，裹在外面的膜，感觉起来像黑洞的膜，似乎该变得越来越轻。当3维的球面坍缩成一点时，相应的黑洞——请坐稳了！——也就没有质量了。尽管听起来太离奇——世上哪有没有质量的黑洞？——我们很快会把它跟我们熟悉的弦物理联系起来。

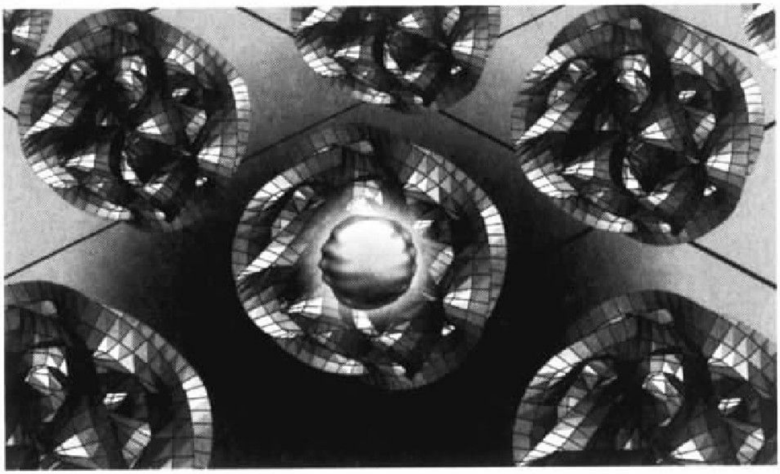


图13.4　膜裹在卷缩维中的球面上时，看起来就像展开维里的一个黑洞。

我们要说的第二点是卡-丘空间的孔洞数目，我们在第9章讨论过，它决定着低能量（从而也是低质量）弦振动模式的数目，而那些振动模式有可能解释表1.1的粒子和力荷。由于空间破裂的锥形变换改变了孔洞的数目（例如，在图13.3中，面包圈的洞被空间的破裂-修补过程消去了），我们希望低质量的振动模式数目也会发生改变。实际上，莫里森、斯特罗明戈和我在具体研究这一点时发现，当新生的2维球面取代卷缩的卡-丘空间里的3维球面时，无质量的弦振动模式恰好增加了1个。（图13.3中面包圈变成沙滩皮球的例子可能会让你觉得模式数应该减少——因为洞少了，但这是低维类比带来的误会。）

为把上面讲的两点结合起来，我们想象一系列卡-丘空间镜头，在这个系列中，一个3维球面正变得越来越小。我们的第1点发现意味着，裹在3维球面上的一张3-膜——在我们看来像一个黑洞——也将越变越小，最后在坍缩的终点变得没有质量。不过，我们还是要问，这是什么意思？借助第2点发现，答案就清楚了。我们的研究表明，那个从空间破裂的锥形变换中新生成的无质量弦振动模式就是黑洞转

化成的无质量粒子的微观图景。于是我们发现，当卡-丘空间经过空间破裂锥形变换时，原来的大质量黑洞会越来越轻，最后转化为一个没有质量的粒子——如零质量的光子——在弦理论中，那不是别的，就是一根以某种特别方式振动的弦。这样，弦理论第一次明确地在黑洞和基本粒子间建立起了直接具体而且在定量上无懈可击的联系。

黑洞“消融”

我们发现的黑洞与基本粒子的这种联系，很像我们早就在日常生活中熟悉的一种现象，物理学叫它相变。相变的一个简单例子是我们在上一章谈到过的：水能以固体（冰）形式存在，也能以液体（液态水）和气体（蒸气）形式存在。这些都是水的相，从一种形式转换为另一种形式，就是相变。莫里森、斯特罗明戈和我证明，这种相变与卡-丘空间从一种形式到另一种形式的空间破裂锥形变换，存在着密切的数学和物理学的相似。从没见过冰的人，不会一下子认识它跟水原来是同一样东西的两个不同的相；物理学家以前也没发现我们研究的黑洞跟基本粒子原来是同一弦物质的不同相。环境的温度决定水以哪种相的形式存在，类似的，卡-丘空间的拓扑形式——即空间形态——决定着弦理论中的某些物理结构是以黑洞还是以基本粒子的形态表现出来。就是说，在第一种相，即原来的卡-丘形态（类似于冰的相），我们看到有黑洞存在；在第二种相，第二种卡-丘形态（类似于液态的水），黑洞发生了相变——可以说它“消融”了——成为基本的弦振动模式。经过锥形变换的空间破裂，将我们从一个卡-丘空间的相引向另一个相。在这个过程中，我们看到黑洞与基本粒子像冰和水那样，是同一枚硬币的两面。我们看到，黑洞“安然”走进了弦理论的框架。

我们有意用同一个水的例子来比喻那些剧烈的空间破裂变换和5个弦理论形式间的转换（第12章），因为后两者有着深刻的联系。回想一下，我们用图12.11来说明5个弦理论是相互对偶的，从而它们统一在一个宏大的理论体系下面。但是，假如我们让额外的维随便卷缩成某个卡-丘形态，那些理论还能自由地从一种图景转换到另一种吗？

——我们还能从图12.11上的任何一点出发达到别的点吗？在根本的拓扑改变结果发现以前，人们认为答案是否定的，因为不知道有什么办法让卡-丘形态连续地从一种转变成另一种。但是现在我们看到，答案是肯定的：通过那些在物理上可能的空间破裂的锥形变换，我们能将任何一个卡-丘空间连续地变成另一个。通过改变耦合常数和卷缩的卡-丘空间几何，我们看到所有的弦结构也都是同一理论的不同相。即使所有额外的空间维都卷缩起来，图12.11的统一也是不可动摇的。

黑洞熵

多年来，一些卓有成就的物理学家都考虑过空间破裂和黑洞与基本粒子相关的可能，尽管这些猜想当初听起来像科幻小说，但弦理论的发现和它融合广义相对论与量子力学的能力，使我们可以将那些可能性推向科学前沿的边缘。这样的成功激励我们进一步追问：我们宇宙的其他一些几十年没能解决的奇妙性质，是不是也将在弦理论的威力面前“屈膝投降”呢？其中最重要的概念是黑洞熵。这是弦理论大显神通的舞台，它成功解决了困惑人们四分之一世纪的一个极深刻重要的问题。

熵是无序和随机的量度。例如，桌上高高地堆着些打开的书、没读完的文章、旧报纸和旧邮件，它就处于一种高度无序的，即高熵的状态。反过来，如果文章按字母顺序堆成一摞，报纸按日期一张张放好，书照作者名字顺序排列，笔放在笔架上，那么，你的书桌就处于一种高度有序的，也就是低熵的状态。这个例子说明了熵的大概意思。但物理学家对熵有一个完全定量的定义，使我们可以用确定的数值来描述一种事物的熵：数值越大，熵越大；数值越小，熵越小。具体说来有点复杂，简单地说，表示熵的数就是一个物理系统的组成元素在不影响整体表现情况下的所有可能组合方式的数量。当书桌整洁的时候，任何一点新安排——如改变书报或文章的堆放顺序，从笔架上拿一支笔——都会干扰原来高度有序的组合。这说明原来的熵很低。反过来说，如果桌子上本来很乱，随便你把报刊文章或邮件怎么

翻动，它还是那么乱，整体没有受到干扰。这说明原来有很高的熵。

当然，我们说重新安排桌上的书报文章，决定哪些安排“不影响整体表现”，是缺乏科学精确性的。熵的严格定义实际上包括数或者计算一个物理系统基本物质组成的微观量子力学性质的可能组合，它们不会影响整体的宏观性质（如能量和压力）。细节并不重要，我们只需要认识熵是精确度量物理系统总体无序性的一个完全量化的量子力学概念。

1970年，贝肯斯坦（Jacob Bekenstein）还在普林斯顿跟惠勒读研究生，他有一个大胆的建议。他的惊人的思想说：黑洞可能有熵，而且量很大。贝肯斯坦的动力来自古老而久经考验的热力学第二定律，这个定律宣告系统的熵总是增大的：事物都朝着更加无序的状态演化。即使你清理好混乱的书桌，减少它的熵，总熵——包括你身体的和房间里空气的——实际上还是增加了。原来，清理书桌时，你得消耗能量；你必须打乱体内脂肪的某些分子次序才可能生成肌肉需要的能量。而当你清理的时候，身体会散发热量，它会激发周围的空气分子进入更混乱的活动和更高的无序状态。所有这些效应都考虑进来，在补充书桌减小的熵之后还有多余的，因而总熵增大了。

贝肯斯坦的问题大概是，假如我们在黑洞事件视界的附近清理好书桌，并用真空泵把房间里新扰动的空气分子抽出来注入黑洞内部幽暗的角落，那么结果会怎样呢？我们还可以问得更极端一些：假如把房间里所有的空气、桌上所有的东西甚至连桌子一起都扔进黑洞里去，把你一个人留在冰冷的空荡荡的完全有序的屋子里，结果会怎样呢？显然，房间里的熵肯定减少了，贝肯斯坦认为，能满足第二定律的惟一途径是黑洞也有熵，当物质进来时，它的熵会充分增大，足以抵消我们看到的洞外熵的减少。

实际上，贝肯斯坦可以借助霍金的一个著名结果来加强他的猜想。霍金曾证明，黑洞事件视界——回想一下，那是遮蔽黑洞的一

个“不归面”，落下去的东西永远也回不来了——的面积在任何物理相互作用下总是增大的。霍金证明，假如一颗小行星落进一个黑洞，或者附近恒星的表面气体被吸积到黑洞，或者两个黑洞碰撞在一起结合成一个……在所有这些过程中，黑洞事件视界的总面积总是增大的。对贝肯斯坦来说，永远朝着更大面积的方向演化与热力学第二定律说的永远朝着更大的熵的方向演化应该有着联系。他指出，黑洞事件视界的面积为它的熵提供了精确的度量。

然而，经仔细考察，多数物理学家认为贝肯斯坦的思想不可能是正确的，原因有两点。第一，黑洞似乎本该是整个宇宙中最有序、最有组织的事物。我们测量黑洞的质量、携带的力荷和它的自旋，就准确地确定了它的一切。凭这样几个确定的特征，黑洞显然没有足够的结构造成无序，就像桌上只有一本书、一支铅笔，随便怎么弄也混乱不起来——黑洞那么简单，哪儿来的无序呢？贝肯斯坦的建议难以理解的第二个原因是，我们刚才讲过，熵是量子力学概念，而黑洞依然生在对立的广义相对论的原野里。在20世纪70年代初，没有什么办法结合广义相对论与量子力学，讨论黑洞可能的熵至少会令人感到不安。

黑洞有多黑

如我们所看到的，霍金也想过他的黑洞面积增大定律与熵增定律间的相似，但他认为那不过是一种巧合，没有别的意思。因为，在霍金看来，根据他的面积增加定律和他与巴丁（James Bar-deen）和卡特尔以前发现的一些结果，假如当真承认了黑洞定律与热力学第二定律的相似性，我们不仅需要把黑洞事件视界的面积当作黑洞的熵，还得为黑洞赋予一定的温度（它的准确数值由黑洞在事件视界的引力场强度来决定）。但是，假如黑洞具有非零的温度——不论多小——那么，根据最基本可靠的物理学原理，它必然发出辐射，就像一根发热的铁棒。然而，谁都知道，黑洞是黑的，不会发出任何东西。霍金和差不多所有的人都认为，这一点绝对排除了贝肯斯坦的建议。另一方

面，霍金更愿意相信，如果有熵的物质被扔进黑洞，那么熵也失去了，这样不是更清楚更简单吗？热力学第二定律，也在这儿终结了。

不过，1974年，霍金发现了真正令人觉得奇怪的事情。他宣布，黑洞并不完全是黑的。如果不考虑量子力学，只根据经典广义相对论的定律，那么像大约60年前发现的那样，黑洞当然不会允许任何事物——包括光——逃出它的引力的掌握。但量子力学的考虑会极大改变这样的结论。虽然霍金也没有广义相对论的量子力学体系，但他还是凭着两个理论的部分结合得到了有限然而可靠的结果。他发现的最重要结果是，黑洞确实在以量子力学的方式发出辐射。

霍金的计算冗长而艰难，但基本思想却很简单。我们知道，不确定性原理使虚空的空间也充满了沸腾和疯狂的虚粒子，它们在瞬间产生，然后在瞬间湮灭。这种紧张的量子行为也出现在黑洞事件视界周围的空间区域。然而霍金发现，黑洞的引力可以将能量注入虚光子，就是说能把两个粒子远远分开，使其中一个落进黑洞。光子对的一个伙伴落进黑洞深渊，另一个光子失去了湮灭的伙伴。霍金证明，剩下的那个光子将从黑洞的引力获得能量和动力，在伙伴落向黑洞时，它却飞离黑洞。霍金发现，从遥远的安全地方看着黑洞的人会看到，虚光子对分裂的最终结果是从黑洞发射出一个光子。这样的过程反反复复在黑洞视界的周围发生，从而形成一股不断的辐射流。黑洞发光了。

另外，霍金还能计算黑洞的温度——从远处的观测者看，即与辐射相应的温度——它由黑洞视界处的引力场强度决定，那还是黑洞物理学的定律与热力学定律之间的相似性所要求的。[3]贝肯斯坦对了：霍金的结果说明应该认真对待那种相似。实际上，这些结果说明那不仅仅是一种相似——本来就是同一样东西。黑洞有熵。黑洞也有温度。黑洞物理学的引力定律不过是热力学定律在极端奇异的引力背景下的另一种表达形式。这是霍金1974年的惊人发现。

现在我们来看那些量有多大。当我们仔细研究了所有细节，可以发现，质量约为3个太阳质量的黑洞，温度大约比绝对零度高一亿分之一度，不是零，但小得可怜；黑洞不黑，但一点儿也不亮。遗憾的是，这样低温的辐射太微弱了，不可能在实验中探测出来。不过，也有例外。霍金的计算还说明，黑洞质量越小，温度越高，从而辐射越强。例如：1颗小行星质量的黑洞会产生大约1万吨氢原子弹的辐射，辐射主要集中在电磁波的$\gamma$射线部分。天文学家在夜空寻找过这些辐射，但除了少数希望渺茫的可能性而外，什么也没找到。这似乎意味着，那样低质量的黑洞即使有也是罕见的。[4]正如霍金常开玩笑说的，这太糟糕了。如果哪天找到了他预言的黑洞辐射，他肯定会得诺贝尔奖。[71]

大质量黑洞的温度在百万分之一度以下，而熵却与它相反。例如，我们计算3个太阳质量的黑洞的熵，结果大得惊人：$10^{78}$，1后面跟78个零！质量越大，熵越大。霍金的成功计算确凿地证明黑洞包含的无序是多么巨大！

但那是什么的无序呢？我们已经看到，黑洞看起来是特别简单的东西，那么惊人的无序是从哪儿产生出来的？关于这个问题，霍金的计算什么也没说。他那广义相对论与量子力学的部分结合可以用来计算黑洞熵的数值，却不能解释它的微观意义。20多年里，一些大物理学家曾试着去认识黑洞的哪些微观性质可能解释熵的意义。但是，在没能将量子力学与广义相对论完全可信地结合起来之前，虽然能看到答案的一点儿影子，谜却还藏在背后。

走进弦理论

那谜直到1996年才揭开。那年，斯特罗明戈和瓦法在苏斯金和森以前发现的基础上，向物理学电子档案发了一篇文章，题目是“贝肯斯坦-霍金熵的微观起源”。在这篇文章里，他们用弦理论认定了某一类黑洞的微观组成，准确计算了相应的熵。他们的研究依赖于一种新

发现的方法，它部分超越了80年代和90年代初的微扰近似。他们的结果完全符合贝肯斯坦和霍金的预言，终于完成了20多年前没能画完的图。

斯特罗明戈和瓦法集中考虑了一类所谓的极端黑洞。这是一些带荷（你也可认为是电荷）的黑洞，而且有与荷相应的可能最小的质量组成。从这个定义，你可以看出它们与第12章讨论的BPS态是密切相关的。实际上，斯特罗明戈和瓦法彻底研究过两者的相似性。他们证明，他们能从一个特别的（具有一定维数的）BPS膜出发构造——当然是在理论上——某些极端的黑洞，并照准确的数学蓝图将它们结合在一起。我们知道构造原子——当然，还是在理论上——可以从一堆夸克开始，将它们组合成质子和中子，然后在周围安排一些沿轨道运动的电子；同样，斯特罗明戈和瓦法证明，弦理论中新发现的物质基元如何能以类似的方法结合起来形成特别的黑洞。

实际上，黑洞是星体演化的一种终结产物。恒星经过几十亿年的核聚变燃尽它所有的核燃料后，就不再有力量——向外的压力——抵抗强大的向内的引力。在不同的条件下，这都将导致恒星巨大质量的灾难性坍缩；它在自身重量下坍缩，最后形成黑洞。斯特罗明戈和瓦法没有用这样现实的方法来生成黑洞，他们说的是“设计者”的黑洞。他们改变了黑洞的形成法则；他们凭着理论家的想象，仔细地、慢慢地、一点一点地将从第二次超弦革命中涌现出来的高维膜缝合在一起，系统地构造了需要的黑洞。

这种方法的力量很快就显露出来了。在完全由理论决定的黑洞微观构造的前提下，斯特罗明戈和瓦法可以很容易地直接计算黑洞微观构成在不影响整体可观测性质（质量和力荷）时的组合方式。然后，他们可以拿组合方式的数目与黑洞视界的面积——即贝肯斯坦和霍金预言的熵——进行比较，他们发现，结果是完全相符的。至少对那类极端的黑洞，斯特罗明戈和瓦法成功运用弦理论准确解释了它们的微观组成和相应的熵。一个困惑人们四分之一世纪的问题就这样解决

了。5

许多弦理论家把这一成功看作支持弦理论的一个重要而令人信服的证据。我们对弦理论的认识还太浅，不可能与实验观测（如夸克和电子的质量）建立直接准确的联系。但我们现在看到，弦理论为发现多年的一种黑洞性质提供了第一个基本解释，那是物理学家多年来用传统方法一直没能解决的。另一方面，黑洞的这个性质紧密联系着霍金关于黑洞辐射的预言，而那个预言在原则上是能够通过实验来观测的。当然，这需要我们在天空确定地找到黑洞，然后构造足够灵敏的仪器来探测它发出的辐射。如果黑洞质量足够小，后一步凭今天的技术是容易实现的。即使实验计划还没有成功，它也再一次强调了弦理论与关于自然世界的确定性物理结果之间的鸿沟是能够填补的。甚至格拉肖这位20世纪80年代以来一直反对弦理论的物理学家最近也说："当弦理论家谈论黑洞时，他们几乎就是在谈可观测现象——那是令人惊奇的。"[72]

黑洞的未解之谜

即使取得了那些令人惊喜的进步，黑洞还有两个百年老问题。一个是关于黑洞对决定论概念的冲击。19世纪初，法国数学家拉普拉斯（Pierre-Simon de Laplace）提出了在牛顿运动定律下像时钟一样运行的宇宙所能带来的最严格也走得最远的结果。

理性能认识某一时刻所有令自然洋溢生机的力和组成其存在物的各自状态，而且，假如理性足够强大，将那些数据投入分析，那么它能将一切运动，从宇宙中最大的物体到最小的原子，都包罗在同一个公式里。对这样的理性来说，没有什么不确定的东西，将来与过去一样，它都看得见。[73]

换句话说，如果知道宇宙每个粒子在某一时刻的位置和速度，我们就可以用牛顿运动定律——至少在原则上——来确定它们在过去或

未来任何时刻的位置和速度。从这样的观点看，一切事物的发生，从太阳的形成到耶稣钉上十字架，到你的眼睛读过这一行文字，都严格遵从大爆炸瞬间过后宇宙各种粒子组成的位置和速度。这种严格的一步步展开的宇宙观跳出了令人困惑的关于自由意志的各种哲学泥潭，但它的重要性却因量子力学的发现而大大消减了。我们看到，海森堡的不确定性原理根本否决了拉普拉斯的决定论，因为从根本上说我们不可能知道宇宙组成物质的准确位置和速度。相反，量子波函数取代了那些经典的性质，它只能告诉我们某个粒子在这里或那里，有这样或那样的速度。

然而，拉普拉斯观的破灭并没有让决定论的思想彻底失败。波函数——量子力学的几率波——的演化仍然遵从准确的数学法则，如薛定谔方程（或者它更准确的伙伴，狄拉克方程和克莱茵-戈登方程）。这告诉我们，量子决定论取代了拉普拉斯的经典决定论：宇宙基本组成在某一时刻的波函数的信息能让“足够强大的”理性去决定以前或未来任何时刻的波函数。量子决定论告诉我们，任何特别事件在未来某一时刻发生的几率完全取决于以前任何时刻的波函数知识。量子力学的几率观极大地弱化了拉普拉斯的决定论，它将“注定的结果”变成“可能的结果”，不过在传统的量子理论框架下，那“可能”还是被完全决定了的。

1976年，霍金宣布，即使这个弱化的决定论也因黑洞的存在而被破坏了。背后的计算当然还是很困难，但基本思想却相当简单。当事物落进黑洞时，它的波函数也跟着被吸收了。这意味着在寻求未来所有时刻波函数的过程中，我们“足够强大的”理性也难免会迷失。为完整地预言未来的波函数，我们需要完全了解今天的波函数。但是如果有些波函数陷入了黑洞的深渊，它们包含的信息也跟着丢失了。

乍看起来，黑洞带来的麻烦似乎不值得忧虑。因为黑洞事件视界背后的一切事物都从我们的宇宙“分离出去了”，我们忽略这些不幸的“坠落者”，有什么不妥吗？另外，从哲学上讲，我们似乎满可以告

诉自己，宇宙并没失去那些落入黑洞的物质所携带的信息，它们不过被锁进了一个我们理性的生命不愿面对的空间区域。在霍金发现黑洞并不全黑以前，这种说法当然没有问题。但霍金向全世界说，黑洞会辐射，事情就不同了。辐射携带着能量，所以黑洞在辐射时会慢慢减小质量——慢慢地“蒸发”。这样，从黑洞中心到事件视界的距离会慢慢收缩；当这遮蔽黑洞的外衣收缩时，原来从宇宙中分离出去的部分空间又能回到我们的宇宙舞台中来。于是，我们的哲学考虑必须面对这样一个问题：被黑洞吞没的事物所携带的信息——我们想象隐藏在黑洞内部的那些数据资料——会因黑洞蒸发而重新出现吗？量子决定论的成立需要这些失去的信息，所以，这个问题深入到了另一个问题的核心：黑洞是否给我们宇宙的演化带来了越来越深层的偶然性因素？

关于这个问题的答案，我写这本书时物理学家还众说纷纭。多年来，霍金强烈主张那些信息不会再出现——黑洞破坏了信息，“在普通量子理论的不确定性上又给物理学增添了新的层次的不确定性。”[74]实际上，霍金和加州理工学院的索恩（Kip Thorne）跟同在那儿的普雷斯基尔（John Preskill）就黑洞所获信息的问题打过赌。霍金和索恩认为那些信息永远地消失了，而普雷斯基尔则站在相反的立场，主张那些信息在黑洞辐射和收缩时还能找回来。赌注呢？还是“信息”：“输家向赢家提供一部赢家选择的百科全书。”

赌局还没分出输赢，不过霍金最近承认，根据我们上面讨论的弦理论对黑洞的新认识，那些信息有可能找到一条重新出现的路径。[75]新看法是这样的：对斯特罗明戈和瓦法研究的（后来许多物理学家也跟着研究过）那类黑洞来说，信息可以储藏在高维膜里，并能从那里还原。斯特罗明戈最近说，这个发现“使许多弦理论家忍不住要欢呼胜利了——信息在黑洞蒸发时回来了。在我看来，结论还下得太早。为弄清这是否正确，我们还有许多事情要做。”[76]瓦法也同意，他说，他“不懂这个问题——哪种情况都是可能的。”[77]回答这个问题是当前

研究的中心目标。如霍金讲的，多数物理学家都愿意相信那信息不会丢失，因为这样能使世界安宁，可以预言。但我相信，如果认真对待爱因斯坦的广义相对论，我们一定允许另外的可能：时空本身打成结，而信息消失在结中。决定信息是否真会丢失，是今天理论物理学的一个主要问题。78

第二个未解之谜是关于黑洞中心点的时空本性。[6]直接像史瓦西1916年那样应用广义相对论，可以证明挤压在黑洞中心的巨大质量和能量将导致时空结构产生吞噬一切的裂隙，卷曲成一种无限曲率的状态——陷入一个时空奇点。根据这一点，物理学家可以得出的一个结论是，因为所有穿过事件视界的物质都注定要落向黑洞中心，而那里的物质没有未来，所以时间本身也在黑洞中心走到了尽头。还有些物理学家，这些年来用爱因斯坦方程探索了黑洞中心的性质。他们发现一个近乎疯狂的结果：黑洞中心可能隐约地联着另一个宇宙的入口。大概地说，我们宇宙时间在哪里结束，相连的另一个宇宙的时间就从哪里开始。

在下一章里我们会讨论这些令人难以想象的结果有什么意义。不过现在我们只谈重要的一点。我们必须记住关键的一课：在极端的大质量、小尺度下，密度大得难以想象，从而不能单独考虑爱因斯坦的经典理论，还必须同时考虑量子力学。这将我们引向一个问题：关于黑洞中心的时空奇性，弦理论有什么说法呢？这也是目前正在研究的一个课题，跟信息丢失问题一样，还没有答案。弦理论灵巧地处理过另外形式的奇异性——我们在第11章和本章第一部分讨论过的那些空间破裂。[7]但认识一种奇性并不意味着认识别的奇性。宇宙的结构能以不同的方式产生破裂。弦理论为某些奇性带来了深刻的认识，但另一些奇性，如黑洞的奇点，至今还躲在弦理论之外。根本的原因还是弦理论太依赖于微扰的工具，在这个问题上，那些近似的方法弱化了我们的能力，从而不能可靠而完全地分析发生在黑洞中心的事情。

然而，随着最近非微扰方法的巨大进步和它们在黑洞其他方面的

成功应用，弦理论家满怀信心地希望能在不远的将来揭开黑洞中心的秘密。

注释

1.专业的读者会发现，在镜像对称下，卡-丘空间里坍缩的3维球面将映射为镜像卡-丘空间里坍缩的2维球面——显然，这使我们回到第11章讨论过的空间翻转情形。然而，两者的差别在于，镜像表述会导致反对称张量场Bμν——镜像卡-丘空间上的复化Kähler形式的实部——为零，这比11章讨论的奇性更加奇异。

2.更准确地说，我们有极端黑洞的例子：具有一定力荷的极小质量组成的黑洞，如第12章的BPS态。类似的黑洞在下面黑洞熵的讨论中还将发挥重要作用。

3.黑洞发出的辐射应该认为跟一个火炉的辐射相同——这是我们在第4章开头讨论的问题，曾在量子力学的发展中起着关键作用。

4.后来发现，由于这些黑洞牵涉到空间破裂锥形变换，不论质量变得多小，它们都不产生霍金辐射。

5.斯特罗明戈和瓦法在他们的初始计算中发现，5维（而非4维）展开的时空能使数学变得更容易。奇怪的是，在完成了5维黑洞熵的计算以后，他们发现，还没有哪个理论家在5维广义相对论的框架下构造过这种假想的极端黑洞。因为只有把结果与这类假想黑洞的事件视界面积进行对比，才能证实他们的计算，所以斯特罗明戈和瓦法开始在数学上构造一个5维黑洞。他们成功了。然后，他们可以很简单地证明，根据弦理论的微观计算得到的熵与根据霍金黑洞事件视界面积预言的熵，应该是完全一致的。注意下面的事实是很有趣的：黑洞的计算是后做的，所以斯特罗明戈和瓦法在做熵的计算时并不知道他们要找的答案。自他们的研究以来，许多研究者，特别是普林斯顿的物理学家Curtis Callan，成功地将熵计算推广到了我们更熟悉的4维时

空的情形，而且所有结果都跟霍金的预言一致。

6.这跟信息丢失问题也有点关系；有些物理学家近年来曾猜想，在黑洞的中心也许有某一“小团”东西藏着那些落入黑洞视界的物质所携带的信息。

7.实际上，本章讨论的空间破裂锥形变换就涉及黑洞，从而应该与它们的奇性问题有关。但是我们记得，锥形破裂只在黑洞所有质量都“脱落”以后才会发生，所以它并不直接与黑洞的奇性问题相关。

## 第14章　宇宙学的沉思

人类自古以来就渴望认识宇宙的起源。也许没有哪个问题像这样超越文化和时代的分隔，它唤起祖先的想象，也引发今天宇宙学家的沉思。从深层说，人们渴望解释为什么会有一个宇宙，它是如何成为我们看到的那个样子的，它是因为什么——什么原理——而演化的。令人惊喜的是，人类今天看到一个正在显露的、能科学地回答那些问题的框架。

今天人们接受的宇宙创生的科学理论认为，宇宙在最初的瞬间经历过最极端的条件——巨大的能量、极高的温度和极大的密度。就我们今天的认识，这些条件把量子力学和引力都牵引到一起了，宇宙的诞生从而成为超弦理论尽情表现的大舞台。我们马上要来讨论那些新奇的发现，但我们还是先回顾一下弦理论以前的宇宙学，也就是人们常说的宇宙学标准模型。

### 宇宙学标准模型

宇宙起源的现代理论应追溯到爱因斯坦完成广义相对论15年以后。尽管爱因斯坦不相信他自己理论的表面意思，不相信它竟包含着一个既不永恒也非静态的宇宙，但弗里德曼相信。如我们在第3章讨论的，弗里德曼发现了我们现在说的爱因斯坦方程的大爆炸解——他

声称宇宙是从一个无限压缩的状态爆炸出来的，现在还处于那场原始大爆炸引起的膨胀中。爱因斯坦坚信他的理论不会有这样随时间演化的解，他发表一篇短文说发现了弗里德曼的结果有个致命的毛病。不过，8个月后，弗里德曼说服了爱因斯坦，他的东西确实没有毛病；爱因斯坦公开认了错，但有点儿漫不经心的样子。不管怎么说，我们可以清楚地感到，爱因斯坦认为弗里德曼的结果跟宇宙没有一点儿联系。但是5年以后，哈勃用威尔逊山天文台的254厘米望远镜详细观测了几十个星系，证明宇宙确实在膨胀着。弗里德曼的结果后来由物理学家罗伯逊（Howard Robertson）和沃克（Arthur Walker）写成更系统更有效的形式，至今还是现代宇宙学的基础。

让我们把宇宙起源的现代理论说得更详细一点儿。大约150亿年以前，宇宙所有的空间和物质从一次奇异的大能量事件爆发出来。（你用不着去寻找大爆炸发生在什么地方，因为它发生在你今天所在的地方，也发生在任何别的地方；我们今天看来分离的不同地方，在宇宙开始的时候都是同一个地方。）大爆炸过后$10^{-43}$秒，即所谓的普朗克时间，计算的宇宙温度大约是$10^{32}$开尔文（K），比太阳内部的温度还高10亿亿亿（$10^{25}$）倍。然后，宇宙随时间膨胀、冷却，在这个过程中，均匀炽热的原初宇宙等离子体开始聚集成团，形成旋涡。大约十万分之一秒后，物质已变得足够冷了（大约10万亿开——比太阳内部温度高100万倍），夸克可以3个成团地聚在一起，形成质子和中子。百分之一秒后，周期表里最轻的一些元素的核也够条件从冷却的粒子等离子体中凝结出来。在接下来的3分钟里，宇宙逐渐冷却到10亿开，出现最多的核是氢和氦，同时也带着些氘（“重”氢）和锂。这就是所谓的原初核合成时期。

接下来的几十万年没发生什么特别的事情，宇宙还是在膨胀、冷却。但是，当温度降到几千开时，汹涌的电子流慢慢流向原子核（多数是氢和氦），原子核捕获住它们，第一次形成电中性的原子。这是重要的一刻：大体上说，从这一时刻开始，宇宙变得透明了。在电子

捕获这一幕之前，宇宙充满了带电的等离子体——有些带正电，如原子核；有些带负电，如电子。只与带电体发生相互作用的光子，落在深深的带电粒子的汪洋里，不停歇地碰撞挤压，要么被偏转，要么被吸收，几乎穿越不了多少距离。因为带电粒子的屏障作用，光子不能自由运动，所以宇宙几乎完全是不透明的，就像在经历浓雾弥漫的早晨，或者遮天蔽日的沙尘暴。但是，当带负电的电子走进带正电的核的轨道，生成电中性的原子以后，带电的屏障消失了，浓雾散开了。从那时起，来自大爆炸的光子就无阻碍地漫游，整个宇宙也慢慢清澈明亮了。

约10亿年以后，宇宙已基本从沸腾的爆发状态安静下来，星系、恒星和行星终于一个个从原初元素的引力束缚堆里产生出来。大爆炸150亿年后的今天，我们也来了，在惊叹宇宙壮丽的同时，也惊讶我们自己能一点点地树起一个合理的而且经得起实验检验的宇宙起源理论。

但是实在说来，我们对大爆炸理论该有几分信赖呢？

大爆炸的检验

用最大的望远镜，天文学家可以在天空看到大爆炸几十亿年后的星系和类星体发出的光，这样他们可以验证那个时期以来的宇宙膨胀，结果都是“真的”。为了检验更早时间的理论，物理学家和天文学家必须用更间接的方法，其中最精妙的一个方法牵涉到所谓的宇宙背景辐射。

你一定给自行车胎打过气，打满气的车胎摸起来有点儿热。打气时耗去的能量有一部分转化来增高车胎里空气的温度。这反映了一个普遍的原理，在很多条件下，被压缩的事物会变热。反过来说，如果什么东西解压了——膨胀了——它就会冷却。空调和冰箱用的也是这个原理。工作物质（如氟利昂）经过循环的压缩、膨胀（同时也蒸发

或凝结），可以让热朝着需要的方向流动。这样地球上寻常简单的物理事实，原来也令人惊奇地发生在整个宇宙。

我们刚才讲过，当电子与核结合成原子以后，光子就自由自在地在整个宇宙中穿行。这意味着宇宙充满了“光子气”，它们沿这样或那样的路径旅行，均匀地洒满宇宙的每个地方。宇宙膨胀时，自由奔流的光子气也跟着膨胀，因为从本质上说，宇宙就是它的一个大容器。一般气体（如轮胎里的空气）的温度在膨胀时会降低，同时，光子气的温度也会随宇宙膨胀而降低。实际上，盖莫夫和他的学生阿菲尔（Ralph Alpher）、赫尔曼（Robert Her-mann）在20世纪50年代，以及迪克（Robert Dicke）和皮贝斯（Jim Peebles）在60年代，就发现我们今天的宇宙应该是一个原始光子的汪洋，它的温度经过150亿年的宇宙膨胀已经冷却到了可怜的绝对零度以上几度。[1]1965年，新泽西贝尔实验室的彭齐亚斯（Arno Penzias）和威尔逊（Robert Wilson）偶然做出了我们时代的一个最重大的发现。他们在寻找无线电通讯干扰的原因时，偶然探测到了大爆炸留下的余温。后来，理论和实验都更精密了，在20世纪90年代初还用美国国家航空航天局（NASA）的“宇宙背景探索者”（COBE）卫星进行了测量。根据这些数据，物理学家和天文学家在很高精度上证实了我们的宇宙确实充满着微波辐射（假如我们的眼睛足够灵敏，就可以看见我们周围的点点微光），温度大约是2.7K，正符合大爆炸理论的预言。具体地说，在宇宙的每个立方米——包括你占据的那个——大约有4亿个光子，它们一起汇成宇宙微波辐射的汪洋，荡漾着宇宙创生的回响。当电视台没有节目时，你看到荧屏上的那些“雪花”，有的就来自大爆炸遗留的暗淡微波。理论与实验的一致，证实了大爆炸之后（ATB）几十万年以来——光子第一次在宇宙自由穿行以来——宇宙演化的图景。

我们对大爆炸理论的检验还能追溯到更早的时间吗？能。通过核理论和热力学的标准原理，物理学家可以很确定地预言在原初核合成阶段（ATB百万分之一秒到几分钟之间）产生的轻元素的相对丰度。

例如，根据理论，宇宙大约23%的元素应该是氦。通过恒星和星云中氦丰度的测量，天文学家获得了令人信服的支持，预言确实是正确的。不过，也许更令人惊讶的是理论关于氘丰度的预言和证实，因为除了大爆炸以外，似乎没有别的天体物理学过程能说明氘的出现——虽然量很小，但宇宙到处都有。这些丰度（以及最近锂丰度）的证实，很好地检验了我们对原初核合成以来的宇宙物理的认识。

这些认识足够我们骄傲了。我们掌握的所有数据都证明，我们的宇宙学理论能描绘宇宙从ATB　0.01秒到150亿年后今天的演化图景。不过，我们还应该看到，新生的宇宙是在瞬息间演化的。我们宇宙100多亿年来持续的特征，在大爆炸之初很短的时间里——比0.01秒还短得多——就第一次深深留下了印迹。所以，物理学家还在往前走，试图弄清更早时期的宇宙。当我们追溯更早的时间，宇宙更小、更热、更紧，于是更迫切需要量子力学来准确描写那时的物质和力。在前面的章节我们看到，点粒子量子场论在点粒子能量一般地处于普朗克能量附近时还能适用。在宇宙学背景下，普朗克能量出现在整个宇宙都压缩在普朗克尺度的一小团时，这时候的密度需要我们发挥想象，找些比喻才能感觉它有多大——在普朗克时刻，宇宙的密度可以用一个字来说，那就是“大”。在这样巨大的能量和密度下，引力论和量子力学不能再像点粒子量子场论那样，看作两个分立的理论。实际上，本书的中心思想就是，在这些高能状态下，我们一定要用弦理论。用现在的话来说，当我们追溯到ATB $10^{-43}$秒（普朗克时间）以下时，会碰到那样的能量和密度，因此，宇宙最早的瞬间原是弦理论的舞台。

我们先还是来看，在标准宇宙学模型中，宇宙从普朗克时间以后到ATB 0.01秒之前都发生了什么。

**从$10^{-43}$秒到0.01秒**

回想一下我们在第7章讲过的（特别是图7.1），在宇宙早期极热

的环境下，引力之外的3种力似乎是结合在一起的。根据这些力的强度随能量和时间而变化的计算，物理学家证明，大约在ATB $10^{-35}$秒以前，强、弱和电磁力原来是一个“大统一”的“超”力。在那种状态，宇宙比它在今天要对称得多。一堆混乱的金属加热熔化后，将形成均匀光滑的液体；同样，我们现在看到的几种力之间的巨大差别，在极早期宇宙的极端高能和高温下也是均匀地融合在一起的。但随着时间的流逝，宇宙不断地膨胀、冷却，量子场论证明，原来的对称性通过许多跳跃的过程丧失殆尽了，最后生成我们今天熟悉的不那么对称的样子。

这种对称性丧失——更准确的说法是对称破缺——背后的物理学是不难理解的。想象一个盛满水的大容器，$H_2O$分子均匀地充满整个容器；不论从哪个角度看，水都是一样的。现在，我们降低温度，看容器会发生什么事情。开始，表面看不出什么。在微观尺度上，也不过是水分子的平均速度减小了。然而，当温度降低到0℃时，你会突然看到激烈的事情发生了。液态的水开始冻结，转化为固态的冰。我们在前一章讲过，这是相变的一个简单例子。就现在的讨论而言，我们需要注意的重要一点是，相变会降低$H_2O$分子所表现的对称性。从任何角度看，液态水看起来都是一样的——显然是旋转对称的——而固态的冰却不是这样的。冰具有晶体的结构，就是说，如果以足够的精度来检验，它跟任何晶体一样，在不同方向有不同的表现。相变使原来的旋转对称性的程度降低了。

尽管我们讨论的只是一个熟悉的例子，结论却是普遍成立的：许多物理系统在温度降低时，在发生相变的地方总会使原来的对称性产生“破缺”。实际上，如果温度改变范围大，一个系统可能会经历一系列的相变。还是看水的例子。如果从100℃以上开始，水是气体，即水蒸气。在这种状态，系统的对称性比在液态时更多，因为这时单个的H2O分子从凝结的液体中解放出来了。它们在容器内四处飞舞，不形成任何小集团，没有哪些分子比别的分子更“亲近”，在高温下，所

有分子都是平等的。如果把温度降到100℃以下，水滴自然在气液相变点凝结出来，而对称性也减少了。继续冷却，经过0℃时，将发生另一次相变，像我们上面说的那样，这一次从液态水到固态冰的相变会再一次大大降低系统的对称性。

物理学家相信，从普朗克时间到ATB 0.01秒，宇宙的行为也像那样，至少经历两次类似的相变。在$10^{28}$K的温度以上，3种非引力作用表现为一种力，具有所有可能的对称性。（在本章末尾，我们将讨论弦理论如何在高温下把引力也包括进来。）但是，当温度冷却到$10^{28}$K以下时，宇宙经历一次相变，3种作用以不同的方式从统一中分离出来；它们的作用强度和方式也开始出现差异。这样，随着宇宙的冷却，在高温下表现的力的对称性就被打破了。不过，格拉肖、萨拉姆和温伯格的研究（第5章）说明，并不是所有的高温对称性都消失了：弱力与电力还密切关联着。随着宇宙进一步膨胀和冷却，在$10^{15}$K的时候——约太阳核心温度的1亿倍——宇宙又经历另一次相变，影响了电磁力与弱力。在这样的温度，两个力还是从以前更对称的统一状态中分离出来，随宇宙不断的冷却而显现出越来越大的差别。两次相变决定了宇宙中作用的3种表现迥然不同的力，即使这样，这一段宇宙的历史回顾也说明那些力实际上是紧密联系在一起的。

宇宙学疑难

普朗克时间以后的宇宙学为我们认识大爆炸瞬间以来的宇宙，提供了一个优美和谐的而且可以计算的框架。不过，跟所有成功的理论一样，新的认识也带来了更多更细的问题。那些问题虽然没有使前面的标准宇宙图景失去意义，还是暴露了某些薄弱的东西，呼唤更深的理论的出现。我们来看其中的一个问题，所谓的视界问题，它是现代宇宙学最重要的问题之一。

宇宙背景辐射的仔细研究表明，不论测量天线对准什么方向，辐

射的温度都是相同的，精确到十万分之一。细想一下会发现，这是很奇怪的事情。在宇宙中相隔那么遥远的地方为什么会有那么一致的温度？我们大概自然会想到，这并不奇怪，因为今天在空中遥遥相对的两个地方，不过是出生以后分离的孪生兄弟，在宇宙最初的瞬间（和任何别的事物一样）本是紧紧相连的。由于它们源自共同的一点，留下相同的痕迹（如温度）也就不足为奇了。

在标准的大爆炸宇宙学里，那种想法是错误的。为什么呢？一碗热汤慢慢冷却到房间的温度，是因为它与周围的冷空气相通。只要等待足够的时间，汤的温度与空气的温度通过相互接触，总会变得相同。但是，如果把汤装在热水瓶里，它会保温很长一段时间，因为与外界几乎没有多少接触。这说明，两个物体的温度趋于相同，是因为它们有长时间稳定的相互交流作用。为了检验刚才说的，现在空间分隔遥远的两点具有相同温度，是因为它们原来曾经接触过，我们必须检验它们在宇宙早期是不是有足够的信息交流。乍看起来，你可能想，那时两点离得很近，交流该是很容易的事情。但空间的邻近只是事情的一个方面，事情的另一方面是时间间隔。

为更完整地考察这一点，我们来看一场宇宙膨胀的“电影”，不过是倒着放的，从今天开始，回到大爆炸的瞬间。因为任何形式的信号和信息的传播速度都以光速为最高极限，所以在某个时刻，空间两个区域的物质，只有在相隔的距离小于光自大爆炸时刻以来能达到的距离，才可能交换热量，从而才可能达到共同的温度。这样，在倒放影片时，我们可看到一场竞争：空间区域离得多近，我们回到过去多远。[2]例如，为了让两个空间位置相距$3\times10^5$千米（即光走1秒经过的距离），我们必须回到ATB　1秒以前，那时候，即使距离那么近，两个空间也不能产生相互影响，因为光需要整整1秒钟的时间才能走过它们之间的距离。2如果空间分离的距离更小，如300千米，我们必须回到ATB　0.001秒以前，刚才的结论也同样成立：两点也不可能产生相互影响，因为在0.001秒之内，光不可能走过300千米的距离。沿着

同样的思路，如果我们的镜头回到ATB $10^{-9}$秒以前，两个空间位置相距30厘米，它们仍然不可能相互影响，因为大爆炸的光没有足够的时间走过那30厘米。这说明，尽管随我们回溯大爆炸，时空间隔会越来越小，但它们未必能像汤和空气那样产生热接触，未必能达到相同的温度。

物理学家已经精确证明了，在标准的大爆炸模型中会产生这个问题。详细计算表明，现在相隔遥远的空间区域没有办法实现能量交换，从而解释不了为什么它们会有相同的温度。物理学家把这个解释不了的宇宙大范围的温度均匀性问题称为“视界问题”——视界在这里说的是我们能看多远；或者也可以说，光能走多远。这个疑难并不意味着标准宇宙模型错了；不过，温度的均匀性确实在强烈提醒我们，宇宙故事里某一幕重要的场景被遗忘了。1979年，物理学家古斯（Alan Guth，现在麻省理工学院）找到了那失去的一幕。

暴胀

视界问题的实质在于，为了让宇宙中任意两个远离的区域靠近，我们必须回到时间的开始。实际上，在那样早的时刻，任何物理影响都不可能有足够的时间从一个区域传到另一个区域。于是，问题就成了，当我们的宇宙影片回放到大爆炸时，宇宙没有足够快地收缩回去。

这只是大概的意思，我们还应该说得更具体一些。视界问题源自这样的事实：膨胀的宇宙像飞出的皮球一样，会因引力的拖曳作用而慢下来。这意味着，为了看到宇宙的两个位置间隔更小，例如，现在距离的一半，我们的宇宙影片必须回放过一半。就是说，为了让那间隔减小一半，我们必须回到大爆炸以来宇宙年龄的一半以前。大致说来，时间越早，两个区域尽管离得更近，但它们的交流越难。

古斯对视界问题的解决现在说起来就很简单了。他发现，爱因斯

坦方程还有另一个解，宇宙在极早期经历过短暂的迅猛膨胀的阶段——在这个阶段里，宇宙空间以意想不到的指数的膨胀速率“暴胀”。指数式的膨胀不像抛向空中的皮球会慢下来，它会越来越快。当我们回放宇宙影片时，迅猛的加速膨胀的镜头表现为迅猛的减速收缩。这意味着为了使宇宙两个位置（在暴胀时期）的间隔减小一半，我们的电影不必回到一半以前——实际上远远用不了那么多时间。这样，两个区域就像热汤和空气那样，有了足够的时间进行热的接触和交换，从而达到相同的温度。

经过古斯的发现和后来林德（Andrei Linde，现在斯坦福大学）、斯坦哈特（Paul Steinhardt）和阿布雷切特（Andreas Albrecht，那时在宾夕法尼亚大学）以及其他许多人的重要修正，标准的宇宙学模型成了暴胀的宇宙学模型。在这个框架下，标准模型在ATB $10^{-36}$秒到$10^{-34}$秒之间的小小“时间窗口”里被修正了——在这个“窗口”里，宇宙膨胀了至少$10^{30}$倍，相比之下，在标准图景中，宇宙在相同时间间隔内只膨胀了大约100倍。这意味着，在ATB $10^{-36}$秒的瞬间，宇宙比它在150亿年以后增大的还多。在暴胀以前，现在相隔遥远的物质离得很近，比在标准模型里近得多，从而可以很容易达到共同的温度。然后，通过古斯的宇宙暴胀——紧跟着标准模型的寻常膨胀——那些空间区域就像我们今天看到的一样，相隔遥远。这样，标准的宇宙学模型经过瞬间暴胀的重要修正，解决了视界问题（以及许多其他我们没有讨论的问题），因而获得了宇宙学家的认同。[3]

我们根据今天的理论，把宇宙从普朗克时间到现在的历史总结在图14.1中。

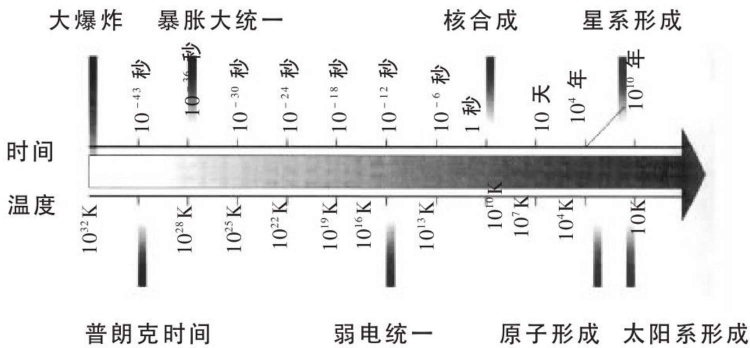



图14.1　宇宙历史上的几个重要时刻。

## 宇宙学和弦理论

在图14.1中，从大爆炸到普朗克时间还留着一丝空白没有讨论。把广义相对论的方程贸然用于这个区域，我们可以发现，当时间越近大爆炸，宇宙会变得越小、越热、越密。在零时间的那一点，宇宙大小消失了，温度和密度顿时成为无穷大，这最明显不过地警告我们，在经典的广义相对论引力框架中树起的宇宙理论模型彻底崩溃了。

大自然坚决地告诉我们，在这样的条件下，我们必须把广义相对论和量子力学结合起来——换句话说，我们必须利用弦理论。目前，弦理论在宇宙学的应用正方兴未艾。微扰论的方法最多能得到大概的轮廓，因为极端的能量、温度和密度需要精确的分析。尽管第二次超弦革命带来了一些非微扰的技术，但它们需要经过一段时间的锤炼才可能满足宇宙学背景下的计算。不过，正如我们现在讨论的，在最近10年左右，物理学家已经迈出了认识弦宇宙学的第一步。下面就是他们发现的一些东西。

弦理论似乎有三条基本途径来修正标准宇宙模型。第一，弦理论以一种今天还不太说得清楚的方式让宇宙有一个可能的最小尺度，这对我们认识大爆炸时刻的宇宙有着重大影响，而标准理论说那时宇宙

收缩到了零尺度。第二，弦理论具有大小半径的对偶性（与它有最小尺度密切相关），我们马上会看到这也有着深刻的宇宙学意义。最后，弦理论具有更多的时空维（大于4），从宇宙学的观点看，我们必须说明所有维的演化。让我们更详细地来讨论这几个问题。

开端有团普朗克尺度的火球

如何用那些弦的理论特征来修正标准宇宙学框架下的结论呢？20世纪80年代末，布兰登伯格和瓦法朝这个方向迈出了重要的第一步。他们得到两点重大发现：第一，当时间倒流，回到开始，温度会不断升高；但当宇宙在所有方向都达到普朗克长度时，温度达到它的最大值，然后开始降低。从直觉说，这一点并不难理解。为简单起见，我们想象（布兰登伯格和瓦法也是那么做的）宇宙所有的空间维都是圆形的。当时间倒流，每一维的半径都会收缩，宇宙的温度也会升高。但是，当每一维的半径坍缩经过普朗克长度时，我们知道，在弦理论中，这在物理上相当于半径从普朗克长度反弹回来。由于宇宙的温度在膨胀中降低，所以可以预料，我们看不到宇宙坍缩到普朗克尺度以下，我们实际只能看到温度在普朗克尺度达到最大，停止升高，然后开始下降。经过仔细计算，布兰登伯格和瓦法证明，事情真是那样的。

这个发现令布兰登伯格和瓦法看到了下面的宇宙学图景。开始时，弦理论的所有空间维都紧紧卷缩成它们最小的可能尺度，大约是普朗克长度。温度高，能量大，但都不是无限的，因为弦理论已经排除了无限压缩的零尺度的起点。在这宇宙开始的瞬间，弦理论的所有空间维都是平等的——完全对称的——都卷缩成一个多维的普朗克尺度的小宇宙。然后，根据布兰登伯格和瓦法的发现，宇宙经历第一次对称破缺；在大约普朗克时间，3个空间维生长出来，而其余的维还保持原来的普朗克尺度。那3个空间维就成了暴胀宇宙图景的主角，它们经历图14.1所概括的普朗克时间以后的演化，膨胀到今天的样子。

为什么是3维呢

紧跟着的一个问题是，什么东西打破了对称性而生出3个膨胀的空间维？就是说，除了我们看到只有3个空间维膨胀到现在的大尺度而外，弦理论是否能提出根本的理由来说明为什么不是其他数目的空间维（如4、5、6等）在膨胀？更对称地讲，为什么不是所有的维都膨胀呢？布兰登伯格和瓦法找到一种可能的解释。回想一下，弦理论的大小半径的对偶性依赖于这样一个事实：当空间维卷缩成圆圈时，弦可以缠绕着它。布兰登伯格和瓦法发现，缠绕着维的弦有限制那个维的倾向，不让它膨胀，就像自行车的外胎套着内胎一样。乍看起来，这似乎在说每个维都会被困住，因为它们都可能被弦缠上。问题是，如果缠绕的弦和它的反伙伴（大概说就是沿反方向缠绕空间维的弦）都考虑进来，它们将立即湮灭，生成一根解开的弦。假如这样的过程发生得足够快、足够多，那么套在空间的许多橡皮套都会解开，那些维也能自由膨胀了。布兰登伯格和瓦法猜想，缠绕的弦只能在3个维上解开，为什么呢？

假定1维直线（如直线王国的空间）上有两个沿同一方向滚动的粒子，如果两个粒子的速度不同，迟早会有一个赶超另一个，从而发生碰撞。不过我们得注意，假如同样两个粒子随机地在2维平面（如平直世界的空间）上滚动，它们很可能永远也不会相遇。第二个空间维为每个粒子打开了一个新路径的世界，那些路径几乎不可能在同一时刻交汇在同一点。在3维、4维或其他更高维的情形，两个粒子就更不容易相遇了。布兰登伯格和瓦法发现，如果把点粒子换成缠绕在空间维上的弦圈，类似的结果也会出现。尽管很难看到，但我们相信，在3个（或更少的）卷缩空间维时，两根缠绕的弦很可能相互碰撞——就像两个点粒子在1维线上运动的情形。但是，在4维或更高维的空间里，缠绕的弦就不太可能发生碰撞——像点粒子在2维或更高维空间一样。[4]

这样，我们看到下面的景象：在宇宙最初的瞬间，源自极高（然

而有限）温度的“骚动”驱使所有卷缩的空间维膨胀，但遇到了缠绕在那些维上的弦的约束，从而它们又回到原来的普朗克尺度的半径。但是，随机的热涨落迟早会使3个空间维长得比别的维大，这样，我们刚才的讨论说明，绕在那3维的弦很可能发生碰撞。大约一半的碰撞牵涉到弦与反弦构成的对，它们将相互湮灭，从而不断地解开约束，使得那3个维能持续膨胀下去。它们长得越大，就越不可能被别的弦所缠绕，因为缠绕大的维度需要更大的能量。这样，膨胀是自我发展的，维长得越大，所受约束就越小。现在我们可以想象那3个空间维如何以上一节讲的方式持续演化，长到我们今天看到的宇宙那么大（或者更大）。

## 宇宙学和卡-丘空间

布兰登伯格和瓦法考虑了一种简单情形，假定所有空间维都卷缩成圆圈。实际上，如我们在第8章看到的，只要这些圆足够大，超越我们今天的观测能力，那它们跟我们看到的宇宙形态就是一致的。但对仍然很小的维来说，更现实的图像是它们卷缩成一个复杂得多的卡-丘空间。当然，问题的关键在于应该是哪一个卡-丘空间？如何决定那个特殊的空间？没人能回答这个问题。但是，结合以前讲过的那些拓扑改变结果和这些宇宙学认识，我们可以提出一个框架。我们现在知道，通过空间破裂锥形变换，任何卡-丘空间都可以演化成别的形式。这样，我们能想象，在大爆炸后喧嚣的热运动中，空间卷缩的卡-丘部分尽管依然很小，却在跳着“热烈的舞蹈”，结构在舞蹈中破裂，破裂后复原，永不停息，历经数不清的不同的卡-丘形态。当宇宙冷却，生出3个大的空间维，卡-丘空间从一种形态向另一种形态转变的脚步也慢下来了，而其余的维度都最终卷缩在某个卡-丘形态，生成我们在周围世界看到的那些物理性质。物理学家面临的挑战是，详尽地认识卡-丘空间的演化，从理论的原则预言它们现在的形态。我们已经看到，卡-丘空间能从一种形态光滑地变成另一种形态，根据这一点，卡-丘形态的选择问题实际上可能归结为一个宇宙学问题。[5]

开始之前

因为没有精确的弦理论方程，布兰登伯格和瓦法在他们的宇宙学研究里做了好多近似和假设。就像瓦法最近说的，我们的工作照亮了一条新途径，弦理论因此可以用来谈一些标准宇宙学方法里的顽固的老问题。例如，我们看到，原初的奇点概念在弦理论中是完全可以避免的。但是，凭我们现在对弦理论的了解，很难在这样极端的条件下做出完全令人信服的计算，所以我们的工作只是投向弦理论宇宙学的第一眼，离最后的结果还远着呢。79

自他们的研究以来，物理学家在深入认识弦宇宙学的路上不断地前进着，走在前头的是维尼齐亚诺和他的伙伴、都灵大学的盖斯佩雷尼（Maurizio Gasperini）。他们提出了自己的一套有趣的弦宇宙学，具有上面讲过的某些特征，但差别也很大。跟布兰登伯格和瓦法的工作一样，他们也靠弦理论的最小长度概念来避免标准的和暴胀的宇宙理论中出现的无限温度和能量密度。不过，他们不认为那意味着宇宙来自一个极热的普朗克尺度的小火球，而认为宇宙可能有一部史前的历史——远在我们所谓的零时间之前就开始了——它将我们引向“普朗克的宇宙萌芽”。

在这大爆炸以前的图景里，宇宙的起点大不同于它在大爆炸框架下的状态。盖斯佩雷尼和维尼齐亚诺的研究告诉我们，宇宙的开端并不是炽热地紧紧卷缩在一起的空间小元胞，而是冰冷的、本质上无限延展的空间。那时候弦理论方程表现出一种迅速的不稳定性——多少有点儿像古斯的暴胀时期——把宇宙的每一点都迅速地驱散开去。他们证明，这使得空间越来越卷曲，温度和能量密度越升越高。6一定时间以后，在大空间里会出现一个毫米大小的三维区域，看起来就像从古斯的暴胀中产生的那个超热超密的小火球。接下来，那个小火球经历寻常大爆炸宇宙学的膨胀，形成我们今天熟悉的宇宙。另外，因为这发生在大爆炸以前的一幕本来就经历了暴胀，所以古斯关于视界问题的答案自然包含在这个“前大爆炸宇宙学”图景里。正如维尼齐亚

诺说的，“弦理论为我们和盘托出了暴胀宇宙学的蓝图。”[80]

超弦宇宙学正在迅速成为活跃而多产的研究舞台。例如，大爆炸之前的图景已经激起了许多热烈而富有成果的争论，我们现在还远不清楚它在弦理论最终将产生的未来宇宙学框架内会起什么样的作用。当然，为了认识这一点，物理学家必须把握第二次超弦革命的方方面面。例如，高维的基本膜的存在会带来什么宇宙学的结果？假如弦理论的耦合常数“偶然”把我们从图12.11的5个边缘引向了中心，我们讨论过的那些宇宙性质会有什么改变吗？就是说，成熟的M理论对宇宙的最初瞬间会产生什么影响？这些核心问题的研究现在正热火朝天。我们已经看到了一线光明。

## M理论与力的融合

在图7.1里我们看到，引力以外的3种相互作用的强度，在宇宙温度足够高的时候是融合在一起的。那么，引力作用的强度如何满足这幅图呢？M理论出现之前，弦理论家可以证明，如果选择最简单的卡-丘空间形态，引力作用差不多也能像图14.2那样与其他3种力融合。弦理论家发现，通过小心选择卡-丘空间形态（当然还有其他一些技巧），可以尽可能避免偏离。但这样事后的调整并不能让物理学家们感到满意。因为现在谁也不知道怎么准确预言卡-丘空间的形态，依靠那些与具体形态细节强烈相关的答案是很危险的。

然而，惠藤证明，第二次超弦革命提供了更强有力的答案。惠藤考察了在弦耦合常数不一定很小的情况下，力的强度会有什么变化。他发现，引力的变化曲线会像图14.2的虚线那样逐渐倾向于与其他力融合，不需要特别选择卡-丘空间形态。尽管为时尚早，但这大概还是说明，在M理论的宏大框架下，宇宙的统一可能会更容易实现。

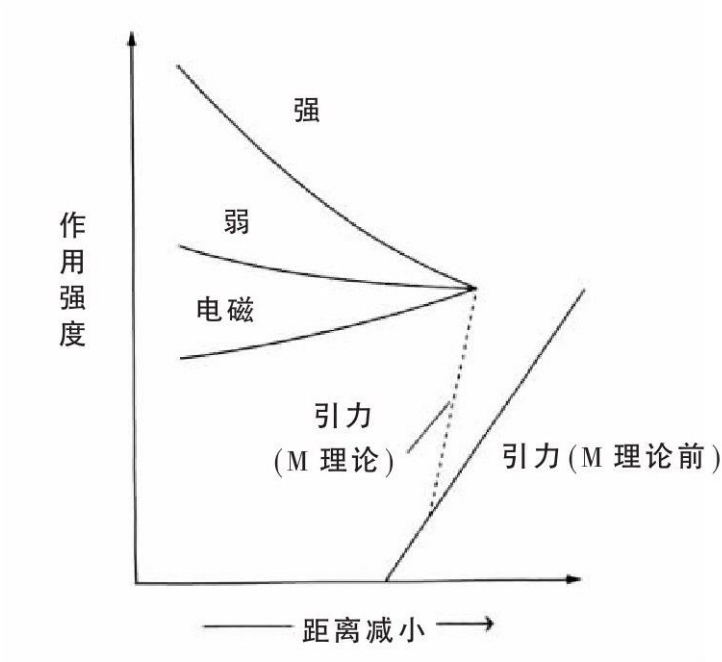



图14.2　在M理论中，4种相互作用自然融合在一起。

这一节和前面几节讨论的发现，是我们朝弦和M理论的宇宙学迈出的头几步，多少还只能说是暂时的结果。在即将到来的年月里，随着弦/M理论非微扰工具的改善，物理学家希望能把它们用于宇宙学问题，并得到某些最深刻的发现。

但我们目前还没有足够有力的方法完全依照弦理论来认识宇宙学，所以我们还是需要一般地考虑宇宙学在寻求未来终极理论的过程中可能发挥怎样的作用。大家应该小心的是，这里的一些思想比以前讨论的更玄，不过它们确实提出了一些未来理论终归要回答的问题。

宇宙学的沉思和终极理论

宇宙学能紧紧抓住我们的心灵，因为认识事物怎么开始，与认识

它们为什么开始，在感觉上是很近的（至少对某些问题是这样）。这并不是说现代科学把“怎么”的问题与“为什么”的问题联结起来了——没有，而且似乎也从来没有谁见过这样的科学联系。但是，宇宙学的研究似乎有希望让我们最完全地认识“为什么”的源头——宇宙的诞生——它至少可以使我们能在一个有科学依据的框架下来提问题。有时候，彻底认识一个问题也就差不多算拥有了问题的答案。

在终极理论的追求中，宇宙学的宏大构思也带来许多更具体的问题。我们相信。宇宙万物今天的表现——即图14.1的时间线上最右端的路线——依赖于物理学的基本定律，但它也可能依赖于宇宙从时间线的左端向右端演化的诸多方面，即使最深远的理论也没能将它们包括进来。

我们不难想象怎么可能是这样的。例如，我们来看皮球抛向天空会发生什么事情。引力定律决定着皮球后来的运动，但我们不能根据那些定律预言球一定会落在什么地方。我们还需要知道球离开我们的速度——包括大小和方向。就是说，我们必须知道球运动的初始条件。同样，还有些宇宙特征具有历史的偶然性——为什么这儿有颗恒星，那儿有颗行星？它们都依赖于一系列复杂的事件，从原则上讲，我们可以追溯宇宙在开始的时候所具有的某些特征。但是，即使最基本的特征，哪怕是最基本的物质和力的粒子的性质，也可能直接依赖于宇宙演化的历史——而演化本身也偶然地依赖于宇宙的初始条件。

实际上，我们在弦理论中已经看到了这种思想的可能体现。随着炽热的早期宇宙的演化，额外的空间维可能从一种形态变换为另一种形态，最后当宇宙冷却下来时卷缩成某个特殊的卡-丘形态。通过这最后的卡-丘形态对粒子质量和力的性质的影响，我们看到，宇宙初始的演化和状态会极大地影响我们今天看到的物理。

我们不知道宇宙的初始条件是什么，也不知道该用什么思想、概念和语言来描绘它们。我们相信，标准和暴胀的宇宙学模型里出现的

那个无限大能量密度和温度的奇异的初始状态，只不过说明那些理论失败了，没能正确描写实际存在的物理条件。弦理论有一点进步，它告诉我们如何避免这种无限的极端；但是，关于事物到底是怎样开始的，我们还是一无所知。事实上，我们的无知更加可怕：我们甚至不知道决定初始条件的问题问得是否合理，不知道这个问题是不是永远超越了任何一个具体的理论——就像要广义相对论来回答把球扔向天空需要多大气力。霍金和加利福尼亚大学的哈特（James Hartle）等物理学家曾大胆尝试把宇宙初始条件的问题带进物理学理论的保护伞下，但他们的努力还没有结果。在弦/M理论的情形，我们今天的认识还肤浅得很，不能决定"包罗万象"的理论候选者是否真的名副其实，不能决定它自己的宇宙学初始条件，当然也就不能把它提到物理学定律的高度。这是未来研究的一个基本问题。

不过，即使不谈初始条件和它们对后来宇宙曲折演化历程的影响，最近的一些猜想仍然意味着任何一个所谓最后的理论都存在解释能力的极限。谁也不知道这些想法是否正确，它们目前当然还处在主流科学的边缘。不过，它们还是以某种方式——尽管存在争论和猜想——让我们看到了未来的终极理论可能会遇到什么样的麻烦。

这个思想源自下面的可能：我们所谓的那个宇宙实际上只是巨大天空的一个小部分，汪洋里无数宇宙岛中的一个。尽管听起来很牵强——最后也许是那样——但林德还是提出一个具体的生成那个大宇宙的机制。他发现，我们以前讨论的短暂而重要的暴胀可能不是惟一的一次事件。他指出，发生暴胀的条件可以多次出现在宇宙众多的独立区域，然后那些区域各自暴胀，演化成为新的分离的宇宙。在每个这样的宇宙中，同样继续着那些过程，新的宇宙又从旧的广大区域里喷涌而来，从而形成一张无穷的宇宙膨胀的大网。这些词儿听起来有点儿累，我们还是用一个流行的词，把这个推广的概念叫多重宇宙，它的每一个组成部分还是叫宇宙。

我们在第7章讲过，我们所了解的一切说明物理学在我们的宇宙

中是和谐的，处处一致的，但这与其他宇宙的物理学没有关系——只要它们与我们是独立的，或者至少离得太远，它们的光还没来得及赶到。所以，我们可以想象物理学是随宇宙的不同而改变的。在某些宇宙，区别可能不太大，例如，电子质量或强力的强度可能比我们的宇宙大（或者小）十万分之一；在另一些宇宙，区别可能很显著，上夸克的质量可能比我们测量的大10倍，电磁力的强度也可能比我们的强10倍，它们同时也给星体和生命带来巨大的影响（如我们在第1章讲的）。还有些宇宙，物理学的差别可能更惊人。例如，基本粒子和力的名单可能跟我们的完全不同；拿弦理论来说，展开的维数也可能不同。紧缩的宇宙可能只有一两个甚至没有展开的空间维，而开放的宇宙可能有八九个甚至10个展开的空间维。如果让我们自由想象，那么定律本身也可能是各不相同的。可能性是无限多的。

问题是这样的。例如我们浏览一下那么多的宇宙，绝大多数都不具备生命存在的条件——至少不会有我们所认识的那些类型的生命。对我们熟悉的物理巨变来说，这是很清楚的：如果宇宙真像花园的水管那样，我们所理解的生命就不会存在。即使不那么剧烈的物理变化，也会影响星体的形成。例如，可能不会有合成复杂生命原子的宇宙大熔炉——像碳、氧等分子，通常都是从超新星的爆发中喷洒出来的。生命的存在离不开具体的物理，从这点看，如果现在问，为什么自然的力和粒子具有我们看到的那些性质，可能有人会回答说：在整个多重宇宙中，那些性质是变化无常的；它们在不同的宇宙可能不同，实际上也的确不同。我们所看到的粒子和力的性质之所以特殊，显然在于它们允许生命的形成。而生命，特别是智慧生命，却是发问的主人：为什么我们的宇宙像这个样子呢？通俗地讲，宇宙万物之所以这样，是因为如果它们不那样，就不会有我们在这儿注意它们。举一个轮盘赌的例子。赢家会惊喜自己能继续赌下去，但他很快就会平静下来。他发现，如果自己没赢，就不可能有那种感觉。多重宇宙的假说也能使我们安静一些，别总想着去解释我们的宇宙为什么会是那样的。

这一路论证不过是一个老思想，有名的人存原理。正如我们看到的，它与我们那个严格的完全能预言的统一理论的梦想是针锋相对的。我们曾经梦想，事情所以那样是因为宇宙不可能是另外一样。多重宇宙不是诗，其中的万物也不像在诗里那么天衣无缝地和谐；它和人存原理一样，描绘了一个无限的宇宙集合，对数不清的变化似乎贪得无厌。多重宇宙的图景是否正确，对我们来说，即使能够理解，也是非常困难的。即使存在别的宇宙，我们也可以想象永远不跟它们往来。不过，多重宇宙的概念扩大了我们的“外面的世界”——相比之下，哈勃发现的银河系外更多的星系就显得太小了——至少会提醒我们：我们对终极理论的要求是不是太多了。

我们应该要求我们的终极理论能给出一幅和谐的描述所有力和物质的量子力学图景。我们应该要求我们的终极理论能给出一个我们宇宙的宇宙论。然而，假如多重宇宙的图景是对的——当然，这是大大的“假如”——那么，要我们的理论来解释粒子质量、电荷和力的具体性质，可能还是要求太多了。

但是必须强调，即使我们接受多重宇宙的设想，也并不一定能说它会损害我们的预言能力。原因呢，简单说来就是，假如我们驰骋想象去考虑一个多重宇宙，我们也会摆脱理论的束缚，去寻找克服多重宇宙那显然的随机性。从相对保守的思想看，我们可以想象，如果多重宇宙的图景是对的，我们能够将我们的终极理论推广到整个宇宙，那个“推广的终极理论”可能会准确地告诉我们，基本的参数为什么那样“洒落”在每一个宇宙？它们是如何洒落下来的？

更激进的思想来自宾夕法尼亚州立大学的斯莫林（Lee Smolin），他从大爆炸和黑洞中心的条件的相似——同样都是挤压在一起大密度的物质——得到灵感，提出每一个黑洞都是一粒新宇宙的种子，新宇宙从种子爆发出来，但永远藏在黑洞视界的背后，我们看不见。斯莫林不仅提出了一种新的生成多重宇宙的机制，还引进来一种新的精神——一种宇宙的基因突变观——把与人存原理相关的科学极

限问题引向尽头。[81]他说，我们来想想看，当一个宇宙从黑洞中心喷出来时，它的物理属性，如粒子质量和力的强度，跟产生它的母宇宙是接近的，但不会完全相同。因为黑洞来自不同星体，而星体的形成完全依赖于粒子质量和作用强度的精确数值，所以，任何一个宇宙能生成多少黑洞，也完全取决于那些参数。于是，“后代”宇宙小小的参数变化可能会比母宇宙更有利于黑洞的形成，从而可能拥有更多的自己的“后代”。[7]这样，经过许多代以后，孕育了很好的黑洞生成条件的子孙宇宙将在多重宇宙中占绝大多数。于是我们看到，斯莫林没有借人存原理，而是提出了一个动力学的机制，说明一代代的宇宙如何一步步接近特殊的参数值——那是最有利于黑洞生成的参数值。

这条思路引出另一种方法，即使在多重宇宙的背景下，它也能解释基本物质和力的参数。假如斯莫林的理论是正确的，假如我们不过是长大的多重宇宙中的一个代表（当然，这些都是“假如”，在许多方面还大有争议），那么，我们测量的粒子和力的参数，应该最有利于黑洞的产生。就是说，我们宇宙的那些参数的一丁点儿改变，都会使黑洞不容易形成。物理学家已经在考察这个预言了，目前还没有大家都能接受的看法。不过，即使证明斯莫林的具体观点错了，它也确实提供了终极理论可能具有的另一种形式。乍看起来，终极理论似乎立场不够坚定，我们可以看到它能描写好多宇宙，而多数都跟我们所在的宇宙无关。另外，我们可以想象那些宇宙都是能够在物理上实现的，从而产生一个多重的大宇宙——表面看，它将永远限制我们的预言能力。然而，实际上这种讨论说明，最终的解释总是可以找到的，只要我们不仅把握了终极的定律，而且还懂得它们在宇宙的大尺度演化的意义。

当然，弦理论和M理论的宇宙学意义在进入21世纪以后都将是一个重大的研究领域。没有能产生普朗克尺度能量的加速器，我们将不得不越来越依赖于大爆炸的宇宙加速器，依赖于它留给我们的遍布宇宙的遗迹，拿它们来当我们的实验数据。凭运气和毅力，我们总有一

天能回答那些基本的问题：宇宙是怎么开始的？它为什么演化成我们看到的苍天和大地？当然，在我们和这些基本问题的完整答案之间，还隔着一大片荒漠。但是，引力的量子理论经过超弦理论的发展，为我们带来了信心和希望。我们相信自己现在掌握了应有的理论工具，可以迈步踏进那片无知的荒漠，经历艰辛万苦之后，我们一定能带着某些最深沉的问题的答案，重新走出来。

注释

1.更准确地说，宇宙应该充满从理想吸收体（热力学叫它“黑体”）发出的那个温度范围的热辐射的光子。霍金曾经证明，黑洞的量子力学辐射也是同样的谱——跟普朗克当年的热炉的辐射一样。

2.这里的讨论传达了所说问题的精神，不过我们淡化了某些微妙的特征，它们与光在膨胀宇宙中的运动有关，也将影响具体的数值。特别的一点是，尽管狭义相对论断言没有比光更快的东西，但这并不排除空间结构所携带的两个光子可能以超光速的速度相互远离。例如，在宇宙第一次透明的大约ATB 30万年的时候，宇宙中分隔90万光年的两个地方还可能相互影响着，虽然它们的间隔超过了光走30万年的距离。那3倍的差距是空间结构的膨胀来补偿的。这意味着，当我们的宇宙大影片回到ATB 30万年时，天空中的两点只要不远离90万光年，就有可能影响彼此的温度。这些具体的数字不会改变我们所说问题的定性特征。

3.关于暴胀宇宙模型和它揭示的问题的详细而生动的描写，请看Alan Guth, The Inflationary Universe（Reading, Mass：Addison-Wesley，1997）。

4.为对数学感兴趣的读者多说几句。这里的结论的根据是，假如两个物体扫过的时空维数的和大于或者等于它们运动所在的时空维数，那么它们总会相交的。例如，点粒子扫过1维时空路径——两个粒子的路径的维数和就是2。直线的时空维数也是2，所以两个粒子总

会相遇（只要它们的速度没有调得来完全相等）。同样，弦扫过2维路径（世界叶），两根弦的和为4，这说明在4维时空（3维空间和1维时间）运动的弦都会相交的。

5.随着M理论的发现和第11维的认识，弦理论家已经开始研究所有7个额外维度的卷缩形式——在那种形式下，每一维差不多都是平等的。这种7维流形的可能选择是所谓的乔伊斯流形，是以牛津大学的乔伊斯（Domenic Joyce）的名字命名的，他第一个发现了那类流形的数学构造技术。

6.专业的读者会发现，我们是在所谓弦参照系中进行描述的，在这样的参照系里，大爆炸之前的曲率增长源自（膨胀驱动的）引力作用强度的增大。在所谓爱因斯坦参照系里，这样的演化表现为加速收缩的一幕。

7.举例来说，在弦理论中，卷缩维的形态从宇宙到“后代”的小小改变，就可能导致那样的演化。根据空间破裂锥形变换的结果，我们知道这样的一个足够长的小变化序列能从一个卡-丘形态经过任何其他的卡-丘形态，让多重的宇宙经历以弦理论为基础的数不清的宇宙的诞生。当多重宇宙历尽新生以后，斯莫林的假说将为我们带来一个期望：我们将拥有一个典型的宇宙，它的卡-丘空间孕育着无限生机。

# 第五篇
# 21世纪的统一

## 第15章　远望

百年以后，超弦理论（或者它在M理论中的角色）该是什么样子，今天恐怕走在最前头的研究者们也看不出来。当我们继续追寻终极理论的时候，在通往更宏大的宇宙蓝图的路上，我们可能会发现弦理论不过是万里长征的一步，我们还会遇到以前从未见过的不同的思想和概念。这一科学历程告诉我们，当我们自以为懂得了自然的一切时，它总还藏着些惊奇，我们只有极大地（有时还得从根本上）改变我们认识世界的思维路线，才可能发现它们。当然，我们还是可以满怀信心地认为——也有人曾那样天真地想象——我们生活在人类历史的一个转折点，宇宙的终极规律将在我们的时代出现。正如惠藤讲的，我觉得我们离弦理论很近了——在我最乐观的时候——我想会有那么一天，理论的最终形式会从天上掉下来落在谁的头上。不过更现实地讲，我觉得我们今天正在构造一个比以往任何东西都更深刻的理论，这个过程将延续到21世纪，那时我就太老了，不可能还有什么有用的思想；年轻的物理学家将去决定，我们是不是真找到了最后的理论。[82]

尽管我们还能感受到第二次超弦革命带来的震撼，还在欣赏它带来的新奇壮丽的图画，但多数弦理论家都认为，可能还要经历第三次、第四次那样的理论革命，才能彻底解放弦理论的力量，确立它作为终极理论的地位。我们已经看到，弦理论打开了一幅宇宙活动的新图画，但还有许多重大的困难和细节需要21世纪的弦理论家用心去思

索。所以，在这最后一章，我们不可能讲完人类追求宇宙最后定律的故事，因为我们还在追求着。我们将把眼光投向弦理论的未来，讨论5个重要的问题——在继续追求终极理论的路上，弦理论家们总会遇到它们的。

弦的基本原理是什么

过去百年里，我们明白了一个大道理，那就是物理学定律总联系着对称性。狭义相对论的基础是相对性原理所赋予的对称性——即常速运动的观测者之间的对称性。表现在广义相对论的引力的基础是等效原理——相对性原理向所有观测者（不论他们的运动状态有多复杂的推广）。另外，强力、弱力和电磁力的基础是更加抽象的规范对称性。

我们讲过，物理学家想把这些对称性树为雄踞在理论的中央的一切解释的基座。从这个观点看，引力的存在是为了让所有的观测者有完全平等的立场——也就是让等效原理能够成立。同样，非引力的存在是为了大自然能遵从它们相应的规范对称性。当然，这种观点不过把力为什么存在的问题转换成为自然为什么遵从相关对称性原理的问题。这肯定也是一个进步，特别是，在有些时候，对称性是自然而然的。例如，为什么一个观测者的参照系需要与众不同的对待呢？更自然的观点显然是，宇宙的规律认为所有观测者的观点都是平等的；这一点通过等效原理和为宇宙结构带来引力而实现了。在引力以外的其他3种力的背后，规范对称性也有同样存在的理由，不过，那需要一些数学背景才能完全理解（如我们第5章讲的）。

弦理论把我们引向了更深的解释的深谷，因为所有这些对称原理——包括那个超对称——都是从它的结构中涌现出来的。实际上，假如历史不像它走过的样子——假如物理学家在百年以前就发现了弦理论——我们可以想象，这些对称性原理都可以通过研究弦理论的性质而发现。但是别忘了，等效原理告诉我们为什么存在引力，规范对称

告诉我们为什么存在非引力，而在弦理论背景下，这些对称都是结果；虽然它们的重要性不容否定，但总归是一个更宏大的理论结构的最终产物的一部分。

等效原理不可避免地带来了广义相对论，规范对称引出了引力以外的3种力，那么，弦理论本身是不是什么更大原理的必然结果呢？那原理可能但不一定是对称性原理。上面的那一段讨论使这个问题显得更尖锐了。写到这里时，还没人能对问题的答案有一丁点儿的认识。为理解它的重要性，我们只需要想想，假如爱因斯坦当年在建立广义相对论时，没有他1907年在伯尔尼专利局的那个把他引向等效原理的“快乐思想”，结果会怎样呢？当然，没有那一点灵感，广义相对论未必就建立不起来，但那一定是异常艰难的。等效原理为分析引力提供了一个简单系统的、强有力的、有条理的框架。例如，我们在第3章对广义相对论的描述主要就依赖于等效原理，而它在理论的数学体系中的作用就更重要了。

弦理论家今天的处境就有点儿像失去等效原理的爱因斯坦。自维尼齐亚诺1968年那独具洞察的猜想以来，弦理论在一点点发现、一次次革命中发展起来了。但是，我们还没有一个组织原理能把所有的发现和理论特征都纳入一个宏大而系统的框架——一个能绝对不可避免地生成每一样基本要素的框架。发现那个原理应该是弦理论发展的重大成果，而它也将以无比的清晰揭示理论深藏的秘密。当然，谁也不能保证真有那样一个基本原理，但百年的物理学进化激励着弦理论家们期待着它的出现。当我们展望弦理论的下一个发展阶段时，寻找那个“能不可避免地带来一切”的原理——整个理论都必然从它喷涌而来——便是头等重要的事情。[1]

## 什么是真正的空间和时间，我们离得了它们吗

在前面的许多章节里，我们自由使用了空间和时间的概念。在第2章，我们讲了爱因斯坦的发现：空间和时间是不可分割的，它们因

一个出人意料的事实而交织在一起，那就是，物体在空间的运动会影响它的时间历程。在第3章，我们通过广义相对论深化了时空在认识宇宙中的作用，看到了时空结构的具体形式如何将引力从一点传递到另一点。第4、第5两章讨论的时空结构的微观量子涨落提出了新理论的需求，将我们引向了弦理论。最后，在接下来的很多章里，我们看到弦理论在宣传宇宙具有的空间维比我们知道的更多，它们有些卷缩成小小的然而复杂的形态，奇妙地经历着空间结构破裂而复原的变换。

我们曾通过图3.4、图3.6和图8.10并借助空间和时空的结构来说明那些思想，那结构仿佛是一片片的物质材料，宇宙就是用它们缝起来的。这些图景有很强的解说力，物理学家常拿它们来作为自己技术研究的直观形象的指南。尽管盯着那些图也能慢慢悟出点儿什么，但我们还是要问，我们所说的宇宙的结构到底是什么意思？

这是一个深远的问题，曾以这样那样的形式提出来，已经争论过几百年了。牛顿宣扬空间和时间是构成宇宙的永恒不变的元素，它原始的结构没有疑问，也不需要解释。他在《原理》（Principia）中写道，“与任何外在物无关的绝对空间，就其本性而言总是保持着相同和不动。与任何外在物无关的绝对的真实的和数学的时间，就其自身和本性而言，总是相同地流逝。”[83]莱布尼兹（Gottfried Leibniz）等人强烈反对，他们声称，空间和时间不过是为了方便概括宇宙中物体与事件间的关系的记录本。一个物体在空间和时间的位置只有通过与其他事物的比较才能显出意义。空间和时间不过是这些关系的词汇，没有别的意思。尽管牛顿的观点在他成功的三大运动定律的支持下统治了200多年，但莱布尼兹的思想（后来得到奥地利物理学家马赫（Ernst Mach）的进一步发展）更接近我们今天的图景。我们已经看到，爱因斯坦的狭义和广义相对论坚决抛弃了绝对和普适的空间和时间的概念。但我们仍然可以追问，在广义相对论和弦理论中演绎着关键角色的时空的几何模型，是否也只是不同位置的空间和时间关系的

方便表达方式呢？或者说，当我们说自己“浸没”在时空结构中时，是不是该认为我们真的浸没在什么东西里呢？

虽然我们在走近一个猜想的领地，弦理论确实能为这个问题提供一个答案。引力子这个最小的引力单元是一种特别的弦振动模式。正如电磁场（如可见光）由无数光子组成一样，引力场由无数引力子组成——就是说，无数根弦在像引力子模式那样振动。另一方面，引力场锁在弯曲的时空结构里，所以，我们自然要将时空结构本身与大量的经历着相同有序的引力子振动模式的弦等同起来。用场的语言说，那么多相同振动的弦的有组织的集合，叫弦的相干态。这是颇富诗意的一幅图画——弦理论的弦成了编织时空结构的丝线——但是应该看到，它的严格意义还有待我们去彻底发现。

不管怎么说，用弦织成的空间结构为我们带来下面的问题。普通的丝织物是在寻常纺织原料上一针一线织出来的。同样，我们可以问自己，时空结构是不是也先有原料底子呢——那该是宇宙结构的一种弦的组合，还没有形成我们认为是时空的组织形式。需要注意的是，我们不太容易准确描绘那种还没织成一个有序整体的一根根振动弦混合在一起的状态，因为从我们寻常的思维方式来说，这预先假定了空间和时间的概念——弦振动所在的空间和它从一刻到下一刻发生形态改变的时间。不过，在那种原始的状态，在形成宇宙结构的弦跳起那整齐相应的舞蹈之前，并没有什么空间和时间。我们的语言还不足以精确把握这些思想，因为事实上那时连以前的概念都没有。总的说来，一根根的弦似乎是空间和时间的“碎片”，只有当它们经过恰当的共振，才可能出现传统的空间和时间的概念。

那样一种没有结构、没有我们所说的空间或时间概念的原始存在状态，可能是大多数人都想象不出来的（我当然也想不出来）。霍金曾说过，摄影师在拍摄黑洞视界的特写镜头时会遇上麻烦，当我们试着构想一个本来是空间和时间的宇宙，而不是以某种方式借用空间和时间概念的宇宙时，也遇到了“范式”的冲突。不过，我们很可能还是

需要同那样的概念打交道，在能完全评价弦理论之前认识它们的作用。原因是，我们现在的弦理论形式预先假定了空间和时间的存在——弦（和在M理论中发现的其他物质基元）在其中往来振动。这样，我们可以在有一个时间维和若干空间维的宇宙中演绎弦理论的物理性质；那些空间维有一定数量是展开的（通常是3个），其余的都卷缩成理论方程所允许的某个空间形态。但是，这有点儿像让一个画家依照数字填颜色，然后根据这个来评价他的艺术创造力。当然，他一定也会在这里或那里表现一些个人的情趣，但凭这样死死限制的作品形式，我们能看出画家有几分才能呢？同样，弦理论的胜利在于它自然融合了量子力学和引力，而引力受空间和时间形式的约束，我们不应该强迫一个理论在已经存在的时空框架里运转。就像应该让画家在空白的画布上开始创作一样，我们应该让弦理论从没有空间和时间的混沌状态开始为自己创造时空的舞台。

我们希望，从“零点”开始——可能是大爆炸以前或者以前的以前的某个时刻（我们只能借时间的词来说，因为没有别的语言工具了）——理论所描写的宇宙将在演化中形成弦相干振动的背景，产生空间和时间的传统概念。这样一个框架如果实现了，将证明空间、时间和相关的维，不是决定宇宙要素的根本；它们不过是从更基本更原始的状态涌出的方便的记号。

M理论的许多方面，经过申克、惠藤、邦克斯、费施勒、苏斯金和其他数不清的人的开拓，已经显露出某个叫零膜的东西——可能是M理论最基本的物质基元，看起来有点儿像大尺度下的点粒子，但在小距离上却有迥然不同的性质——它大概能让我们看一眼没有空间和时间的世界。我们记得，弦告诉我们在普朗克尺度下传统的空间概念失去了意义，他们的研究表明，零膜在本质上也告诉我们相同的结论，而且还为我们打开了一扇小窗，让我们看到一个新的非传统的起主导作用的框架。零膜的研究说明，普通的几何被所谓的非对易几何取代了，那主要是法国数学家康尼斯（Alain Connes）发展起来的一

门数学。[2]在这个几何框架下，传统的空间和距离的概念消失了，我们看到的是迥然不同的概念景观。不过，当我们关心比普朗克长度更大的尺度时，物理学家证明那些传统的空间概念又将重新出现。非对易几何框架离我们期待的那个“零点”的开端大概还很遥远，但它让我们隐约看到，为了包容空间和时间，更复杂的框架可能带来些什么。

为建立一个不借助先存在的空间和时间概念的弦理论寻找正确的数学工具，是弦理论家们面临的最重要的问题之一。如果认识了空间和时间是如何出现的，我们将向下面那个关键问题的答案迈出一大步：到底会出现什么样的几何形式呢？

弦理论会重塑量子力学吗

量子力学的原理以令人惊讶的精度统治着我们的宇宙。即使这样，半个多世纪以来，物理学家在构建理论时所采取的策略，从结构上讲，却把量子力学放在比较次要的位置。在构想一个理论时，物理学家常常从纯经典的语言出发——那是麦克斯韦甚至牛顿时代的物理学家都能完全领会的语言——忽略量子几率、波函数等事物，然后，在经典的框架上添加量子的概念。这种思想方法一点儿也不奇怪，因为它直接反映了我们的科学历程。开始的时候，宇宙看起来是由植根在经典概念的定律统治着的，例如，在一定的时刻一个粒子有一定的位置和速度。当我们做过仔细的微观考察之后，才发现那样的经典思想需要修正。我们发现的历程是从经典框架走向量子关系的框架，物理学家今天还继续走在那条路上，去创建他们的理论。

弦理论当然也是这样走过来的。描写弦理论的数学形式开始是一组描写一根无限细小的经典丝线的运动的方程，大体上说，这样的方程牛顿在300年前就能写出来。后来，这些方程被量子化了。就是说，通过50多年来物理学家们发展起来的一套系统方法，这些经典方程转移到了量子力学的框架，几率、不确定性、量子涨落等概念，都自然包括进来了。实际上，我们在第12章已经看到了这个过程的作

用：圈过程（见图12.6）包含着量子概念——在这种情形，即量子力学生成的瞬间的虚弦对——圈的数目决定着量子力学效应的精度。

从经典的理论图景开始，然后包括量子力学的特征，这种策略多年来取得了丰硕的成果。例如，它是粒子物理学标准模型的基础。但是，这种方法在弦理论和M理论那样远大的蓝图面前可能显得太保守了，而且越来越多的事实说明它很可能真是软弱无力的。原因是，我们既然认识了宇宙受量子力学原理的支配，我们的理论从一开始就应该是量子力学的。多年来，我们从经典图景出发取得了一次又一次的成功，那只是因为我们还没有追到宇宙的最深处，在那样的深度，过去粗略的方法会让我们迷失方向。但是，在弦/M理论的深度上，我们很可能会走到那条经过战斗考验的道路的尽头。

关于这一点，重新考虑从第二次超弦革命涌现出的某些发现（如图12.11总结的那些），我们可以找到具体的证据。我们在第12章讨论过，5个弦理论统一背后的对偶性告诉我们，发生在一个弦理论体系中的物理过程可以用任何其他理论的对偶语言来解释。乍看起来，新的解释似乎跟原来的图景没有什么关系，但事实上对偶性的力量正表现在这里：通过对偶性，同一个物理过程可以用许多迥然不同的方式来描写。这些结果难以捉摸，也令人惊讶，但我们还没有讲它们最重要的特征是什么。

对偶性语言通常是这样转换的：在一种弦理论的描述下，一个过程强烈依赖于量子力学（例如，涉及弦相互作用的过程就不会在经典物理的世界里发生），而从另一个弦理论看，它却稍微与量子力学有些关系（例如，它的具体数值由量子思想决定，而定性形式却跟它在纯经典的世界里一样）。这意味着，量子力学完全交织在弦/M理论基础的对偶对称性中，它们是固有的量子力学对称性，因为有一个对偶的描述是强烈依赖于量子力学考虑的。这有力地说明，弦/M理论的完全实现——从根本上包括新发现的对偶对称性——不能跟传统路线一样从经典开始走向量子化。经典的出发点必然会忽略对偶对称

性，因为它只有在量子力学的考虑下才会表现出来。相反，完全的弦/M理论一定会打破传统模式，而将以一个羽翼丰满的量子力学理论的形式出现在我们面前。

目前，还没有谁知道该怎么做。但许多弦理论家都预言，我们认识上的下一个重大变革是重塑量子原理，将它融入我们关于宇宙的理论。例如，像瓦法说的，“我想，能解决许多疑难的量子力学新体系就躲在角落里。我想，许多人都同意，最近揭示的对偶性为量子力学指出了一个新的几何的方向，在那个几何框架下，空间、时间和量子性质将不可分割地结合在一起。”[84]而照惠藤的说法，“我相信量子力学的逻辑状况即将发生某种方式的改变，就像引力的逻辑状况在爱因斯坦发现等效原理后的改变那样。这个过程对量子力学是远不完全的，但我想人们总有一天会回过头来，把今天看作它的开始。”[85]

我们可以有把握地乐观地想象，一个弦理论框架下的重新树立的量子力学原理将产生一个更有力的理论体系，为我们回答宇宙是如何开始的，为什么会有空间和时间之类的事物——这个体系还将带我们走近莱布尼兹的疑问：为什么会有而不是没有？

弦理论能经受实验的检验吗

在我们以前讨论过的弦理论的特征中，下面3个也许是最重要、最应该牢记的。第一，引力和量子力学是宇宙如何表现的最主要内容，任何一个可能的统一理论都必须包括它们。弦理论实现了这一点。第二，通过物理学家在过去100年的研究，还揭示了其他的重要思想——许多都被实验证实了——它们对我们认识宇宙起着关键作用。举几个例子，这些思想包括，自旋、物质粒子的族结构、信使粒子、规范对称、等效原理、对称破缺和超对称性，等等。所有这些概念都自然出现在弦理论中。第三，在传统理论如标准模型中，有19个可以调整的参数来保证理论与实验测量的一致。弦理论则不同，它没有可调的参数。从原则上讲，它蕴含的一切都是完全确定的——它们

应该提供绝不含糊的检验，以判别理论是对还是错。

从“原则上”的理由走到“实际上”的事实，一路上还有许许多多的障碍。在第9章我们讨论过一些技术上的困难，如决定额外维的形态，现在仍然拦在路中央。在第12章和13章，我们把这些和另外一些拦路石放到了一个更大的背景下——为了更准确地理解弦理论，我们看到，M理论就在那样的背景下出现了。当然，为了完全认识弦理论和M理论，我们不仅需要付出巨大艰辛的劳动，也一样需要天才的发现。

在前进的每一步，弦理论都在寻找而且还将继续寻找能通过实验观测的理论结果。我们大概不会忘记第9章讲的那些未来发现弦理论证据的可能。而且，随着认识的深入，弦理论一定会出现一些难得的过程或特征，为我们提供其他间接的实验信号。

但最引人瞩目的是，通过寻找第9章讨论的超对称伙伴粒子，超对称性的证实应该是弦理论的一个里程碑。我们记得超对称性是在弦理论的理论考察中发现的，也记得它是弦理论的核心部分。它的实验证明，对弦来说尽管是间接的，然而也是诱人的。另外，寻找超对称伙伴粒子也应该是受欢迎的一个挑战，因为如果发现了超对称性，它的意义远不只是回答它是否与我们的世界有关这样的简单问题。超伙伴粒子的质量和力荷将具体揭示超对称性是如何融入自然律的。那样，弦理论家面对的挑战将是，超对称性是否完全可以通过弦理论来实现和解释？当然，我们可以更乐观地希望，在未来的10年——在日内瓦的巨型量子对撞机投入运行以前——弦理论的认识会取得巨大进展，能在发现超对称伙伴粒子之前做出一些关于它们的具体的预言。那么，证实那些预言将是科学史上不朽的一页。

科学的解释有极限吗

解释一切，即使从特定意义说，认识宇宙的力和基本组成的所有

方面，也是科学面临的一个最大挑战。超弦理论第一次为我们提供了一个足以迎接这个挑战的框架。但是，我们真能完全实现理论的承诺，计算出那些量吗——如夸克的质量、电磁力的强度和其他决定宇宙形形色色特征的数值？正如前几节讲的，我们需要克服数不清的障碍才可能达到那些目标——当前的头等大事是建立一个非微扰的弦/M理论体系。

但是，即使我们准确认识了在更新更明晰的量子力学框架下建立起来的弦/M理论，我们仍然可能算不出粒子的质量和力的强度，有这个可能吗？我们可能还得借助于实验测量而不能靠理论计算来获得那些数值，是吗？而且，会不会那样，这些失败不是说我们还需要寻求更深层的理论，而是正好说明这些实在的观测性质本来就没有什么解释？

所有这些问题都是可能的。正如爱因斯坦很多年前讲的，“宇宙最不可理解的事情是它是可以理解的。”[86]在飞速进步的时代，动人的发现很容易使我们盲目信任自己对宇宙的理解力，然而，理解力也许真有它的尽头。也许我们不得不接受这样的事实，当我们达到了最深层的科学认识以后，宇宙依然有一些问题不能解决。也许我们不得不承认，宇宙的有些特征之所以那样，纯粹是因为偶然，因为一个事故，或者因为“魔鬼的选择”。科学在昨天的成功激励着我们去想，只要有足够的时间，巨大的努力总能揭开宇宙的奥秘。但是，遇到科学解释的绝对极限——那不是技术的障碍或趋势，而是人类理解进步的边缘——那可是奇特的事情，过去的经历对今天的我们也无能为力了。

尽管这个问题与我们对终极理论的追求有着重大关系，但我们还解决不了它。实际上，我们以一般方式提出的科学解释极限的问题，可能永远也没有答案。例如，我们已经看到，即使我们关于多重宇宙概念的猜想，乍看起来提出了科学解释的极限，实际上还可以通过幻想别的理论来解决，至少在原则上那个理论能重新找回预言能力。

从这些思考中我们看到了宇宙学在决定一个终极理论时的作用。我们讲过，超弦宇宙学是一个年轻的领域，即使从年轻的弦理论自身的标准说，它也是年轻的。无疑，它将成为未来若干年里的一个基本的研究焦点。随着对弦/M理论性质的新认识，我们能更清楚地判别在统一理论上的那些努力有什么宇宙学意义。当然，那些研究也许有一天会令我们相信，科学解释确实存在着极限。但是，它们也可能预示着一个新时代的到来——那时我们可以宣告，宇宙的基本解释终于找到了。

走向未来

虽然我们的技术把我们限制在地球和它在太阳系的近邻，但依靠思想和实验的力量，我们也在探索空间和外太空。特别是在过去的100年里，经过无数物理学家的努力，自然最深藏的一些秘密都被揭示出来了。这些解释的萌芽一旦破土生长起来，就会在我们原以为了解的世界展现一片新的景象，那壮丽的风光是我们从来不曾想过的。衡量一个物理理论有多深，是看它在多大程度上向以前那些似乎不可改变的世界观提出了严峻的挑战。以这个标准来看，量子力学和相对论的深刻超乎了任何人的想象：波函数、几率、量子隧道、不停歇的真空能量涨落、空间与时间的融合、同时的相对性、时空结构的弯曲、黑洞、大爆炸……谁能想到那个直观的、机械的、像时钟一样运行的牛顿的世界竟显得那么狭小，谁能想到在事物平凡的表面下还藏着一个令人心跳的新世界？

不过，这样一些改变我们思维模式的发现也只是一个更大的包罗万象的历史的一部分。物理学家坚信，不论关于大事物的定律还是小事物的定律，都应该结合成一个和谐的整体，他们怀着这样的坚定信念在孜孜不倦地追寻着隐藏的统一理论。追寻还远没到头，但通过超弦理论和从它演化而来的M理论，一个融合量子力学、广义相对论以及强弱和电磁力的强有力的框架终于出现了。这些进步给人们以前的世界观带来的冲击是巨大的：一圈圈的弦、一颗颗跳动的液滴，把宇

宙生成的万物都统一地归结为形形色色的振动模式，而那些精密的振动所在的宇宙空间具有许多隐藏的维度，能极端地卷缩起来，不停地经历结构的破裂和修复。谁能想到，引力和量子力学会融入一个包罗所有物质和力的统一理论，为我们对宇宙的认识带来那么巨大的革命？

当然，如果我们继续追求更完全的可以计算的超弦理论，一定还有更大的惊奇在等着我们。通过M理论的研究，我们已经看到，在普朗克尺度下隐藏着一个新奇的世界，那里可能没有空间，也没有时间。在另一个尽头，我们也看到，我们的宇宙也许只是在巨大的波涛汹涌的汪洋（即所谓的多重宇宙）表面上无数跳荡的泡沫中的一个。这些思想都是我们今天所能提出的最远的想象，它们可能预示着我们的宇宙认识的下一步该怎么走。

我们一直在放眼未来，期待着潜藏的奇迹；我们也应该回顾过去，走到今天的那段历程同样令人惊讶。追寻宇宙的基本定律是人类的一出独特的戏剧，它解放了思想，丰富了精神。爱因斯坦曾生动描述过他本人对引力的追求经历——“那是在黑暗中焦虑地摸索的年月，满怀着强烈的渴望，有过信心，也有过动摇和疲惫，但最后终于看见了光明。”[87]——这当然也是一切人类奋斗的写照。我们每一个人都在以自己的方式追求真理，渴望知道我们为什么是这样。我们在攀登中发现和解释堆起的大山，每一代人都稳稳站在前辈的肩头，勇敢地走向顶峰。我们的子孙后代会不会有一天站在峰顶上无限清晰地俯看苍茫而壮丽的宇宙，我们不得而知；但每一代人总会向上爬得更高，令人想起布朗诺夫斯基（Jacob Bronowski）的话：“每个时代都有一个转折点，都有一种新的认识和评判世界秩序的方法。”[88]我们这一代人也在惊讶我们自己的新宇宙观——我们认识世界秩序的新方法——实际上也在实现我们自己的价值，把我们搭成人类的阶梯，通向遥远的星辰。

注释

1.有些理论家从全息原理看到了这个思想的一点影子，那个概念最先是苏斯金和著名荷兰物理学家特胡夫特（Gerard’t Hooft）提出来的。我们知道，全息图从特制的2维胶片再现3维图像。苏斯金和特胡夫特提出，我们遇到的所有物理事件都可以完全通过定义在更低维世界的方程来说明。听起来这就像根据人的影子来画肖像，但根据第13章讨论的黑洞熵，我们能领会它的意思，了解苏斯金和特胡夫特的部分动机。回想一下，黑洞的熵决定于事件视界的表面积——而不是视界所包围的空间体积。于是，黑洞的无序和它相应的可能包含的信息都记录在表面积的2维数据里。黑洞的事件视界仿佛就是一幅全息图，它抓住了黑洞内所有3维信息的内容。苏斯金和特胡夫特把这个思想推广到整个宇宙，他们指出，发生在宇宙“内部”的每一件事情都只是定义在遥远边界面上的数据和方程的表现。最近，哈佛的物理学家马尔达西纳（Juan Maldacena）的研究，以及后来惠藤和普林斯顿的物理学家古塞（Steven Gubser）、克里巴诺夫（Igor Klebanov）和波利亚可夫（Alexander Polyakov）的重要工作，证明了至少在一定条件下，弦理论体现着全息原理。看来，弦理论统治下的宇宙的物理学似乎有一个等价的图景，那里只有发生在边界面上的物理——边界的维当然一定比内部的维更低，这是如何实现的，物理学家目前正在积极研究。有些弦理论家提出，彻底认识全息原理和它在弦理论中的作用，将导致第三次超弦革命。

2.假如你熟悉线性代数，你可以有一个简单而且相关的办法来考虑非对易几何，那就是，以矩阵代替传统的笛卡儿坐标。在乘法下，笛卡儿坐标是可以交换的（即对易的），而矩阵是不能交换的（不对易的）。

# 科学名词解释

暴胀宇宙（Inflationary cosmology）对标准的大爆炸宇宙学关于极早期图景的修正，宇宙在极短时间内经历巨大的膨胀。

闭弦（Closed string）形如线圈的一种弦。

波长（Wavelength）波相继两个峰或谷之间的距离。

波函数（Wave function）量子力学赖以建立的几率波。

波粒二相性（Wave-particle duality）量子力学的基本特征，说明物质表现出既像波又像粒子的性质。

玻色弦理论（Bosonic string theory）最早的一种弦理论，包含的振动模式都是玻色子。

玻色子（Boson）一种粒子或弦振动模式，自旋为整数，通常是力的信使粒子。

不确定性原理（Uncertainty principle）海森堡发现的一个量子力学原理，说明宇宙中存在着某些特征，如位置和速度，不可能同时完全精确地认识。微观世界的这种不确定性在我们考虑的距离和时间尺度越小时变得越显著。由于量子不确定性，粒子和场在所有可能的数值间波动、涨落。这说明微观世界是一片沸腾汹涌的量子涨落的汪洋。

缠绕模式（Winding mode）缠绕在卷缩空间维上的弦构形。

缠绕能量（Winding energy）缠绕在卷缩空间维上的弦所具有的能量。

缠绕数（Winding number）弦在卷缩空间维上的缠绕圈数。

场，力场（Field, Force field）从宏观观点看，力通过它来产生作用。由空间每一点的力的大小和方向的集合来表现。

超对称标准模型（Supersymmetric standard model）包含超对称性的粒子物理标准模型的推广，使已知的基本粒子类型增多了一倍。

超对称量子场论（Supersymmetric quantum field theory）融合了超对称性的量子场论。

超对称性（Supersymmetry）联系整数自旋（玻色子）与半整（奇）数自旋粒子（费米子）的一种对称性原理。

超伙伴（Superpartners）由超对称性联系的自旋相差1/2单位的两个粒子。

超微观（Ultramicroscopic）长度小于普朗克长度（从而时间小于普朗克时间）。

超弦理论（Superstring theory）融合了超对称性的弦理论。

超引力（Supergravity）结合广义相对论和超对称性的一种点粒子理论。

虫洞（Wormhole）连接宇宙不同区域的一个类似于管道的区域。

初始条件（Initiation condition）描写物理系统初始状态的数据。

大爆炸（Big bang）现在流行的一种宇宙理论，认为今天膨胀的宇宙来自约150亿年前的一个高能量、高密度的压缩状态。

大收缩（Big crunch）一种假想的未来宇宙状态，今天的膨胀将停止、反转，然后所有空间和物质坍缩到一起；是大爆炸的逆过程。

大统一（Grand unification）把引力以外的所有力融合在一个框架下的一种理论模型。

单圈过程（One-loop process）微扰论计算中的一项，只包括一个虚弦对（在点粒子理论中即一个虚粒子对）。

倒数（Reciprocal）乘积为1的两个数互为倒数。如3的倒数是1/3，2的倒数是1/2。

等效原理（Equivalence principle；Principle of equivalence）广义相对论的核心原则，声称加速运动与在引力场中静止（在足够小的观测区域内）是不可区分的。这是推广了的相对性原理，一切观测者，不论运动状态如何，只要承认引力的存在，都可以说自己是静止的。

第二次超弦革命（Second superstring revolution）约从1995年兴起的一场弦理论发展运动，开始认识了一些理论的非微扰特征。

电磁波（Electromagnetic wave）电磁场中的波扰动，所有这样的波都以光速运动。例如，可见光、X射线、微波、红外线等。

电磁场（Electromagnetic field）电磁力的作用场，由空间每一点的电力线和磁力线构成。

电磁辐射（Electromagnetic radiation）电磁波所携带的能量。

电磁规范对称性（Electromagnetic gauge symmetry）量子电动力学的基本规范对称性。

电子（Electron）带负电荷的粒子，一般出现在原子核外的轨道上。

对称破缺（Symmetry breaking）系统表现的对称性的减少，通常与相变相关。

对称性（Symmetry）物理系统不因某种变换而改变的性质。例如，球是旋转对称的，因为它不因旋转而改变。

对偶，对偶性，对偶对称性（Dual, Duality, Duality symmetry）表面上完全不同的两个或多个理论能得出完全相同的物理结果，它们就是对偶的。

多孔面包圈，多柄圈（Multi-doughnut, Multi-handled doughnut）面包圈（环）的推广，有多个孔或柄。

多维孔洞（Multidimensional hole）面包圈在高维情形的推广。

多重宇宙（Multiverse）假想的扩大的宇宙，我们的宇宙不过是无数独立的不同宇宙中的一个。

2维球面（Two-dimensional sphere）见球面。

翻转变换（Flop transition）空间卡-丘形态的演化形式。结构在演化中破裂，然后自我修复，而产生的结果在弦理论背景下则是温和的、可以接受的。

反粒子（Antiparticle）反物质的粒子。

反物质（Antimatter）与寻常物质有相同的引力性质，但有相反

的电荷和相反的核力荷。

非微扰的（Nonperturbative）有效性独立于近似的、微扰的计算的理论所具有的特征；是理论的一种精确特性。

费曼路径求和（Feynman sum-over-paths）一种量子力学思维形式，假想粒子从一点运动到另一点要经过两点间所有可能的路径。

费米子（Fermion）一种粒子或弦振动模式，具有半奇数的自旋，一般为物质粒子。

辐射（Radiation）波或粒子携带的能量。

干涉模式（Interference pattern）来自不同位置的波因为相互叠加而形成的波动图样。

高维超引力（High-dimensional supergravity）高于4维的超引力理论。

共振（Resonance）物理系统振动的一种自然状态。

观测者（Observer）理想化的（通常是假想的）测量相关物理性质的人或仪器。

光电效应（Photoelectric effect）电子在光照下从金属表面逸出的现象。

光滑，光滑空间（Smooth, Smooth space）空间结构平直或微弱弯曲的特殊区域，没有褶皱、破裂或任何类型的裂痕。

光子（Photon）电磁力场的最小单元，电磁力的信使粒子，最小的一束光。

光子钟（Light clock）通过记数一个光子在两个镜面间往返的次数来测量时间流逝的一种假想时钟。

广义相对论（General relativity）爱因斯坦建立的引力理论，证明了空间和时间通过它们的弯曲传递引力。

规范对称（Gauge symmetry）引力以外的3种力的量子力学图景所依据的对称性原理。它包括物理系统在电荷、位置（空间）和时间改变下的一些不变性。

核（Nucleus）原子的中心，由质子和中子组成。

荷（Charge）见力荷。

黑洞（Black hole）理论预言的天体的最后归宿，巨大的引力场将一切事物（包括光）紧紧捕获在一个极小的空间里（由它的事件视界所包围的空间）。

黑洞熵（Black-hole entropy）黑洞内部所表现的熵。

宏观（Macroscopic）指寻常经历的典型的较大尺度，与微观相对。

环（Torus）面包圈的2维表面。

积（Product）两个数相乘的结果。

极端黑洞（Extremal black hole）具有一定质量下的最大可能力荷的黑洞。

加速度（Acceleration）物体速度大小或方向的改变量，参见速度。

胶子（Gluon）强力场的最小作用单元，强力的信使粒子。

镜像对称性（Mirror symmetry）弦理论背景下的一种对称性，互为镜像的两个卡-丘空间，在选择为弦理论的卷缩维的几何形式时，将生成相同的物理。

卷缩维（Curled-up dimension）没有大的可观测延展的空间维；挤压、卷曲在微小的空间区域，因而不能直接探测。

绝对零度（Absolute zero）自然界可能出现的最低温度，开尔文（Kelvin）温标的零度（0K），或约-273摄氏度（-273℃）。

均匀振动（Uniform vibration）弦的整体运动，没有任何形变。

卡拉比-丘（成桐）空间，卡拉比-丘（成桐）形态（Kalabi-Yau space，Kalabi-Yau shape）弦理论要求的多余空间维所能卷缩形成的空间（形态），与理论的方程相应。

卡鲁扎-克莱茵理论（Kaluza-Klein theory）将多余卷缩维与量子力学结合在一起的一种理论。

开尔文（Kelvin）以绝对零度为基准的一种热力学温标。

开弦（Open string）弦的一种，有两个自由端。

克莱茵-戈登方程（Kline-Golden equation）相对论量子场论的一个基本方程。

夸克（Quark）受强力作用的粒子，有6种类型（上、下、粲、奇、顶、底）和3种“颜色”（红、绿、蓝）。

快子（Tachyon）质量（平方）为负的粒子，它在理论中的出现通常会带来矛盾。

拉普拉斯决定论（Laplacian determinism）宇宙像时钟那样运行，某一时刻宇宙的完整信息能够决定它在未来和过去任意时刻的状态。

黎曼几何（Riemann geometry）描写任意维弯曲形态的数学框架，在爱因斯坦广义相对论的时空描述中起着关键作用。

力荷（Force charge）粒子具有的对某种力的作用产生一定响应的性质。例如，粒子的电荷决定了它对电磁力的反应。

粒子加速器（Particle accelerator）使粒子产生近光速度并将其挤压在一起以探测物质结构的机器。

粒子物理学标准模型（Standard model of particle physics）关于引力外的3种力及其与物质相互作用的高度成功的理论。是量子色动力学与弱电理论的强有力统一。

粒子族（Families）物质粒子有组织地分成3组，每一组被称为一组。每一组粒子带有相同的电荷和核力荷，但后一组的粒子有更大的质量。

量子（Quanta）根据量子力学，物质可以分解成的最小物理单元。例如，光子是电磁场的量子。

量子场论（Quantum field theory）见相对论量子场论。

量子电动力学（Quantum electrodynamics, QED）融合了狭义相对论的电磁力和电子的相对论量子场论。

量子几何（Quantum geometry）为了描写量子效应显著的超微观尺度下的空间的物理而对黎曼几何进行的修正。

量子决定论（Quantum determinism）量子力学的一个性质。一个系统在某一时刻的量子态完全决定了它在过去和未来任意时刻的量子态。然而量子态的知识只能决定它在未来某个时刻实际发生的几率。

量子力学（Quantum mechanics）主宰宇宙的一个理论框架，有许多陌生的基本特征，如不确定性、量子涨落、波粒二相性等，在原子和亚原子的微观尺度上将变得极为显著。

量子泡沫（Quantum foam）见时空泡沫。

量子弱电理论（Quantum electroweak theory）见弱电理论。

量子色动力学（Quantum chromodynamics, QCD）融合了狭义相对论的强力和夸克的相对论量子场论。

量子隧道（Quantum tunneling）量子力学的一个特征，指物体可以通过在牛顿经典物理定律看来不可能通过的势垒。

量子引力（Quantum gravity）能成功融合（可能有一定修正）量子力学和广义相对论的一个理论。弦理论是这种理论的一个例子。

量子幽闭（Quantum claustrophobia）见量子涨落。

量子涨落（Quantum fluctuation）系统在微观尺度上由于不确定性原理而出现的湍流行为。

零维球面（Zero-dimensional sphere）见球面。

洛伦兹收缩（Lorentz contraction）狭义相对论表现的一个特征：运动物体在运动方向上显得缩短了。

麦克斯韦理论，麦克斯韦电磁理论（Maxwell theory, Maxwell

electromagnetic theory）麦克斯韦19世纪80年代在电磁场概念基础上提出的统一电与磁的理论，证明了可见光是电磁波的一种。

膜（Brane）弦理论中出现的任何延展体。1-膜是弦，2-膜即通常的膜，3-膜有3个延展方向（维），等等。一般地说，p-膜有p个空间维。

牛顿万有引力理论（Newton's universal theory of gravity）一种引力理论，宣称两个物体间的相互吸引力正比于物体质量的乘积，反比于物体间距离的平方；后来被爱因斯坦的广义相对论所取代。

牛顿运动定律（Newton's laws of motion）以绝对的不可变易的空间和时间概念为基础的描写物体运动的定律；在爱因斯坦发现狭义相对论之前，这些定律是不可动摇的。

耦合常数（Coupling constant）见弦耦合常数。

频率（Frequency）波在每一秒钟所完成的波动循环数。

平直性（Flat）欧几里得几何应服从的法则。例如完全光滑的桌面以及它的高维推广。

普朗克长度（Planck length）约$10^{-33}$厘米。在小于它的尺度下，时空结构的量子涨落开始变得剧烈。弦理论中一根弦的典型大小。

普朗克常数（Planck constant）记作h，量子力学的基本常数。它决定着能量、质量、自旋等物理量的离散单位的大小，微观世界即照那些单位分离。它的值为$6.62\times10^{-34}$焦·秒。

普朗克能量（Planck energy）约1000千瓦时；探测普朗克长度下的距离所必需的能量。弦理论中一根振动弦的典型能量。

普朗克时间（Planck time）约$10^{-43}$秒。这时的宇宙大小约为普朗克长度；更准确地说，它是光经过普朗克长度的时间。

普朗克张力（Planck tension）约$10^{39}$吨。弦理论的典型张力。

普朗克质量（Planck mass）约质子质量的1000亿亿倍，$10^{-5}$克；大约一粒灰尘的质量。弦理论中一根振动弦的典型等价质量。

奇点（Singularity）时空的空间几何彻底破碎的地方。

强力，强核力（Strong force, Strong nuclear force）4种基本力中最强的，主要作用是把夸克束缚在质子和中子里，并把质子和中子束缚在原子核中。

强力对称性（Strong force symmetry）作为强力基础的一种规范对称性，与物理系统在夸克颜色转移下的不变性相关。

强耦合（Strong coupled）弦耦合常数大于1的理论。

球面（Sphere）球体的外表面。我们熟悉的3维球体的表面是2维的（可以用两个数来表征，如地球表面的“经度”和“纬度”）。不过，球面概念还更一般地用于任何维的球体和球面。1维球面是圆周的更富想象的名字；0维球面是两个点（解释见正文）。3维球面则很难画出来，它是4维球体的表面。

曲率（Curvature）物体、空间或时空形态偏离平直形态，从而也偏离欧几里得几何法则的程度。

热力学（Thermodynamics）19世纪发展起来的理论，描写物理系统的热、功、能、熵及其相互演化。

热力学第二定律（Second law of thermodynamics）关于总熵永

远增大的定律。

人存原理（Anthropic principle）关于宇宙为什么具有我们观测的性质的一种解释原则，假如宇宙不是那样，就可能不会形成生命，从而也就不会有我们来观测那些变化。

弱电理论（Electroweak theory）在一个统一框架下描写弱力和电磁力的相对论量子场论。

弱规范玻色子（Weak gauge boson）弱力场的最小单元，弱力的信使粒子，叫W或Z玻色子。

弱力，弱核力（Weak force, Weak nuclear force）4种基本力之一，最有名的表现是辐射衰变中的力。

弱耦合（Weak coupled）弦耦合常数小于1的理论。

3维球面（Three-dimensional sphere）见球面。

熵（Entropy）物理系统无序的量度；由在不改变系统外在表现的条件下重新安排系统组成的方式数来决定。

11维超引力（Eleven-dimensional supergravity）一个很有希望的高维超引力理论，在20世纪70年代发展，后来被忽略，最近被证明为弦理论一个重要部分。

时间膨胀（Time dilation）狭义相对论中的现象，在运动观测者看来，时间流变慢了。

时空（Spacetime）最先出现在狭义相对论的空间和时间的统一体，可以看作宇宙赖以形成的一种“结构”，为宇宙事件的发生提供了动力学的舞台。

时空泡沫（Spacetime foam）根据传统的点粒子观点，超微观尺度的时空结构所表现的扭曲、混乱特征。弦理论出现以前，这是量子力学与广义相对论互不相容的一个根本原因。

时空破裂翻转变换（Space-tearing flop transition）见翻转变换。

史瓦西解（Schwarzschild solution）广义相对论方程关于球状物质分布的解，这个解的一个结果是可能存在黑洞。

世界叶（World-sheet）弦运动扫过的2维曲面。

视界问题（Horizon problem）与下面事实相关的一个宇宙学难题：宇宙中分隔遥远的区域具有几乎完全相同的性质，如温度。暴胀宇宙模型提供了一个答案。

手征的，手征性（Chiral, Chirality）基本粒子物理学区别左和右的特征，说明宇宙不完全是左右对称的。

速度（Velocity）物体运动矢量，包括运动方向和速率。

统一理论，统一场论（Unified theory, Unified field theory）在一个包罗一切的框架下描写一切力和物质的理论。

拓扑（Topology）几何形态的分类性质，同一类型的不同形态可以不经过任何结构破坏而相互变换。

拓扑差异（Topologically distinct）两种形态如果不经过某种形式的结构破坏不可能从一个变形为另一个。

拓扑改变变换（Topology-changing transition）发生结构破坏的空间演化，从而成为不同拓扑的空间。

微扰方法（Perturbative approach, Perturbative method）见微扰论。

微扰论（Perturbative theory）一种简化问题的方法，先寻找一个近似解，然后将原来忽略的、未系统考虑的细节包括进来。

维（Dimension）空间或时空的一个独立方向或坐标轴。我们周围的空间有3个维（上下、前后、左右），而我们熟悉的时空有4个维（前面的3个空间轴和1个时间轴）。超弦理论要求宇宙有更多的空间维。

无穷大（Infinities）在点粒子框架下的广义相对论和量子力学计算中出现的一类典型的无意义结果。

无质量黑洞（Massless black hole）弦理论中的一类特殊黑洞，它原来可能有巨大质量，但随着空间卡-丘部分的收缩而变得越来越轻，当卡-丘空间收缩到一个点时，它的质量也完全消失了，成为无质量的。在这种状态，它不再表现通常的黑洞性质，如事件视界。

狭义相对论（Special relativity）爱因斯坦关于在无引力作用时的空间和时间定律。（参见广义相对论。）

弦（String）基本的1维物体，是弦理论的物质基元。

弦理论（String theory）一个统一理论，提出自然的基本组成不是0维的点粒子，而是1维的被称为弦的小细丝。弦理论和谐地统一了量子力学和广义相对论这两个已知的然而互不相容的关于“小”和“大”的理论体系。通常是超弦理论的简称。

弦模式（String mode）弦可能表现的形式（如振动模式，缠绕形式等）。

弦耦合常数（String coupling constant）一个（正）数，决定着一根弦如何分裂为两根，或者两根弦如何结合成一根——弦理论的基本过程。每个弦理论都有自己的耦合常数，其值可以由方程来决定；不过，我们现在的方程还不足以得出任何有用的信息。耦合常数小于1意味着微扰论是有效的。

相（Phase）用于物质时，指它可能的状态：固相、液相和气相。更一般地说，指一个物理系统在所依赖的某个性质（温度，弦耦合常数值，时空形式等）发生改变时所可能表现的图像。

相变（Phase transition）物理系统从一个相到另一个相的转变。

相对论量子场论（Relativistic quantum field theory）融合了狭义相对论的关于场（如电磁场）的量子力学理论。

相对性原理（Principle of relativity）狭义相对论的中心原理，声称所有匀速运动的观测者都遵从相同的物理学定律，因每一个匀速运动的观测者都可以说自己是静止的。这个原理被推广为等效原理。

信使粒子（Messenger particle）力场的最小作用单位，力的微观携带者。

虚粒子（Virtual particles）从真空瞬时生成的粒子，根据不确定性原理，依靠借能量而存在，然后在瞬间湮灭，从而还回能量。

薛定谔方程（Schrödinger equation）量子力学中决定几率波演化的方程。

延展维（Extended dimension）大的能直接显现的空间（和时空）维，是我们熟悉的维，与卷缩维相对。

引力（Gravitational force）自然界4种相互作用中最弱的那一

种，曾经用牛顿的万有引力定律来描写，后来用爱因斯坦的广义相对论来描写。

引力子（Graviton）引力场的最小作用单元，引力相互作用的信使粒子。

宇宙微波背景辐射（Cosmic microwave background radiation）在大爆炸产生的随着宇宙膨胀而稀薄、冷却的充满宇宙的微波辐射。

宇宙学常数（Cosmological constant）爱因斯坦为了满足静态宇宙而添加在他原来的广义相对论方程里的一个修正常数，可以解释为真空的常数能量密度。

宇宙学的标准模型（Standard model of cosmology）大爆炸理论加上总结在粒子物理学标准模型的关于引力外的3种力的认识。

原初核合成（Primordial nucleosynthesis）发生在大爆炸后最初3分钟时的原子核生成。

原子（Atom）物质的基本构成要素，由原子核（包括质子和中子）和核外的一群绕核旋转的电子构成。

杂化E弦理论（Heterotic-E string theory）即杂化$E_8 \times E_8$弦理论，5种超弦理论之一，其中的闭弦的右向振动与II型弦的相同，而左向振动涉及玻色弦。与杂化O弦理论有重要而微妙的区别。

杂化O弦理论（Heterotic-O string theory）即杂化O（32）弦理论，5种超弦理论之一，其中的闭弦的右向振动与II型弦的相同，而左向振动涉及玻色弦。与杂化E弦理论有重要而微妙的区别。

振动模式（Oscillatory pattern, Vibrational mode, Vibrational pattern）弦振动时准确的峰、谷数和振幅大小。

振动数（Vibrational number）描写弦的均匀振动的能量的数；这个整体运动的能量不同于与形变相关的能量。

振幅（Amplitude）波峰的最大高度或波谷的最大深度。

质子（Proton）带正电荷的粒子，一般存在于原子核，由3个夸克组成（2个上夸克和1个下夸克）。

中微子（Neutrino）电中性粒子，只服从弱作用。

中子（Neutron）电中性粒子，通常出现在原子核中，由3个夸克组成（2个下夸克，1个上夸克）。

锥形变换（Conifold transition）空间的卡-丘部分的一种演化形式，其结构先发生破裂，然后自我修复，但产生的物理结果在弦理论背景下是温和的、可以接受的。这里出现的破裂比翻转变换的更严重。

自旋（Spin）我们所熟悉的同一名称的量子力学形式。粒子有一定量的内禀自旋，要么是整数，要么是半整数（以普朗克常数为单位），永不改变。

I型弦理论（Type I string theory）5种超弦理论的一种，包括开弦和闭弦。

2-膜（Two-brane）见膜。

IIA型弦理论（TypeIIA string theory）5种超弦理论中的一种，包括具有左右对称振动模式的闭弦。

IIB型弦理论（TypeIIB string theory）5种超弦理论中的一种，包括具有左右对称振动模式的闭弦。

3-膜（Three-brane）见膜。

ATB“大爆炸之后”（“After The Bang”）的缩写，通常用来表示自大爆炸以来的时间。

BPS态（BPS states）超对称理论中的物质构成，其性质可以通过以对称性为基础的论证完全决定。

M理论（M-theory）第二次超弦革命出现的理论，将以前的5个超弦理论统一在一个宏大框架内。M理论似乎是一个包含着11个时空维的理论，还有很多方面有待认识。

T. O.E.（包罗万象的理论，Theory of Everything）囊括所有力和物质的量子力学理论。

W玻色子（W boson）见弱规范玻色子。

Z玻色子（Z boson）见弱规范玻色子。

# 推荐读物

Abbot, Edwin A. Flatland：A Romance of Many Dimensions.Princeton：Princeton Univer-sity Press，1991.

Barrow, John D. Theories of Everything.New York：Fawcett-Columbine，1992.

Bronowski, Jacob. The Ascent of Man.Boston：Little, Brown，1973.

Clark, Ronald W. Einstein, The Life and Times.New York：Avon，1984.

Crease, Robert P. and Charles C.Mann.The Second Creation.New Brunswick, N.J.：Rutgers University Press，1996.

Davies, P. C.W.Superforce.New York：Simon&Schuster，1984.

Davies, P. C.W.and J.Brown, eds.Superstring：A Theory of Everything？Cambridge, Eng.：Cambridge University Press，1988.

Deutsch, David. The Fabric of Reality.New York：Allen Lane，1997.

Einstein, Albert. The Meaning of Relativity.Princeton：Princeton University Press，1988.

Einstein, Albert. Relativity.New York：Crown，1961.

Ferris, Timothy. Coming of Age in the Milky Way.New York：Anchor，1989.

Ferris, Timothy. The Whole Shebang.New York：Viking，1997.

Feynman, Richard. The Character of Physical Law.Cambridge, Mass.：MIT Press，1995.

Gamow, George. Mr.Tompkins in Paperback.Cambridge, Eng.：Cambridge University Press，1993.

Gell-Mann, Murray. The Quark and the Jugar.New York：Freeman，1994.

Glashow, Sheldon. Interactions.New York：Time-Warner Books，1988.

*这些书多数都有中译本，有些已经（或）即将收在我们的《第一推动丛书》里，正文提到的都在脚注中说明了，没提到的，读者也不妨找来看看。——译者

Guth, Alan H. The Inflationary Universe.Reading, Mass.：Addison-Wesley，1997.

Hawking, Stephen. A Brief History of Time.New York：Bantam Books，1988.

Hawking, Stephen, and Roger Penrose. The Nature of Space and Time.Princeton：Prince-ton University Press，1996.

Hey, Tony, and Patric Walters. Einstein's Mirror.Cambridge,

Eng.：Cambridge Univer-sity Press，1997.

Kaku, Michio. Beyond Einstein.New York：Anchor，1987.

Kaku, Michio. Hyperspace.New York：Oxford University Press，1994.

Lederman, Leon, with Dick Teresi. The God Particle.Boston：Houghton Mifflin，1993.

Lindley, David. The End of Physics.NewYork：Basic Books，1993.

Lindley, David. Where Does the Weirdness Go？NewYork：Basic Books，1996.

Overbye, Dennis, Lonely Heart of the Cosmos. New York：HarperCollins，1991.

Paise, Abraham. Subtle Is the Lord：The Science and Life of Albert Einstein.New York：Oxford University Press，1982.

Penrose, Roger. The Emperor's New Mind.Oxford, Eng.：Oxford University Press，1989.

Rees, Martin J. Before the Beginning.Reading, Mass.：Addison-Wesley，1997.

Smolin, Lee. The Life of the Cosmos.New York：Oxford University Press，1997.

Thorne, Kip. Black Hole and Time Warps.New York：Norton，1994.

Weinberg, Steven. The First Three Minutes.New York：Basic Books，1993.

Weinberg, Steven. Dreams of a Final Theory.New York：Pantheon，1992.

Wheeler, John A. A Journey into Gravity and Spacetime.New York：Scientific American Library，1990.

# 译后记

近来，物理学家发现我们真的处在一个很特别的时代，宇宙的三个不同意义的物质密度竟然好像是相等的，它们的巧遇仿佛在告诉我们，我们不能凭自己的知识来决定我们所在的宇宙的命运。这应该是科学大戏最有人情味的结果，它在保留宇宙的自由时，也尊重了科学的自由。奥古斯丁在《忏悔录》里说过，上帝为那些胆敢对宇宙起源说三道四的家伙准备了地狱。而我们今天感觉，大自然是慷慨地任人去评说她的生死的，一点儿不在意，似乎也不屑来否定或者肯定她的儿女好不容易形成的思想。大自然是宽容的，喜欢自我否定的是物理学和物理学家自己。

这本书写的，从某种意义上说就是物理学家自我否定的最新历程。关于宇宙、物质和时空的基本问题，它都问尽了。一个包罗万象的理论，实际上就是问一切似乎不是问题的问题；过去理所当然的一切，现在都成了问题。从前问这些问题，像屈原问天一样，满怀着激情。（尽管现在也能从中发掘某些接近科学的东西，但发问的人那时并不知道。）现在我们问类似的问题，不是因为谁想起了什么古老的歌谣，而是理论容不得那些问题占有那样当然的地位。过去的问题通常是，“事物为什么这样？”而今天我们还要问，“事物为什么不是那样？”物理学家从前以为，即使所有基本的一切都不再基本了，我们还能留下时空的拓扑结构，现在看来，这一点也难留下来了。借作者的话说，超弦理论讲的，就是自爱因斯坦以来的空间和时间的故事。物理学家要从零开始问出需要的东西。惠勒在20多年前的一次演说里讲过一个有趣的故事，可以帮助我们理解这一点：我们来做一个提问游戏，每个人心里原来没有任何预先想到的东西，各自提出任何可以

想到的问题，只要它们的回答不相互矛盾，每一次回答将引出新的问题，这样一直把游戏进行下去，看最后能得到什么确定的东西。

过去我们爱问宇宙的外面是什么。看过这本书以后，我们大概会问，如果“终极理论”找到了，还会有什么问题吗？实际上，这就是那个问题。也许，一切物理问题最终都要回到人类自身来的，因为宇宙的琴弦是我们拨动的。

本书原来的题目很抽象，The Elegant Universe，我不知该用哪个词来形容宇宙，特别是书里描写的那个宇宙。因此，我不得不换一个书名，幸运的是有个现成的名字在那儿：宇宙的琴弦；其实关于弦理论的好多书都会讲这个比喻（本书当然也没例外）；这个名字的另一点意义是把科学与艺术自然地联系起来了。李可染先生为李政道先生画过一幅画，是由线条构成的，题目就是“超弦生万象”。借宗白华先生论素描的话说，“抽象的线条，不存于物，不存于心，却能以它的匀整、流动、回环、曲折，表达万物的体积、形态与生命；更能凭借它的节奏、速度、刚柔、明暗，有如弦上的音、舞中的态，写出心情的灵境而探入物体的诗魂。”那说的不就是我们的弦吗？大雕塑家罗丹（A.Rodin）说得更干脆：“一个规定的线贯通着大宇宙，赋予了一切被创造物……”

这大概是第一本系统讲述弦理论的科普读物，在翻译中自然遇到些困难，特别是不知道某些专门的名词该怎么说（译者没读过几篇有关的中文文献，不知它们有没有约定的译法）。如Heterotic，有说“奇异”的，也有说“杂优”的，前一个说法不太确切，后一个说法意思很准确（原文借生物学名词Heterosis，杂种优势），但不太好听。我以为说“杂化”就蛮好（当然不能说“杂种”喽）。一些数学味道极浓的词，如Conifold，Orbifold，Flop Transition等，我在读过原始论文后还是没能想到直观的表达方式。我请教过作者，怎么用一个普通的词来表达那些空间变换过程。作者的回答跟他在书里讲的一样具体，例如，他建议用Glue-fold来说Orbifold，这在英文里当然是通俗一点

儿了，可惜中文没有这样造词的。最后，译者大胆杜撰了几个自以为不那么佶屈聱牙的说法。好在原文都附在后面，读者自能鉴别。实际上，我以为离开了数学背景，许多概念是不大可能说明白的。甚至，直观和通俗有时竟能成为误会的根源。不过作者讲得很巧妙，读者不会迷失方向的。请读者找那些数学来看看，真的很有趣。

另外，作者为不同文字的译者——可见这书在世界各地都受欢迎——提供了一些他写作时涉及的美国现实生活的背景材料，我在相应的地方注明了，也许能为本书增添几点花絮。

原书只有一个索引，我把其中的人名分出来另列一个，为的是更加醒目，也为着突出今天活跃在弦舞台上的演员们。这些名字绝大多数只能在专业期刊上见到，还没有约定的中文译名，所以都保留原文，如果翻译过来，恐怕熟悉他们的读者会感到陌生，想走近他们的人也不知该去找谁了。（正文里的译名有的很勉强，可能不符合中文译名的通例。根据读音还原的几个中国和日本名字可能也不准确，我向他们和读者说声对不起。）

我曾想把重要的原始文献都列举出来，但那要费很多工夫，而且也超出了一本普及读物的范围。更何况，这是一门开放的学问，就在我译这本书的几个月里，大概又出现了近千篇的文章——遗憾的是在我们身边很难听到超弦的声音，在国内最重要的科学刊物上也少见它的踪影。不知这本书能吸引多少未来的流水高山的知音？

译者

昆明东川，银河影下

2001年7月28日

# 重印后记

本书最后一章关于全息图像的注释，在作者今年出版的一本新书里（The Fabric of the Cosmos：Space, Time, and the Texture of Reality, New York：Alfred A.Knopf，2004），仍然是最后几页讨论的问题："宇宙是全息图吗？"这似乎有点儿奇怪；我觉得，续写"琴弦"，似乎可以从宇宙全息开始写起——具体说，可以从霍金的黑洞信息的丢失写起，或者从Juan Maldacena的一篇文章写起。当然，那新书并不是"续篇"，而是重新整理了我们对时空的一些基本观念，当然主要还是弦带来的。不管怎么说，在过去的几年，弦理论有了些可以宣扬的东西，"对偶性"又添了许多新内容。夸张一点说，对偶无处不在。它们的代表是所谓"AdS/CFT对应"，不明白意思的同学，不妨先把它当作跟TCL、KFC一样的品牌记住，有关的普及读物，也许很快就能看到。

不过，对偶更像天堂的因缘，什么时候能下落凡间，还不知道。而且，80年代对弦理论的批评（如作者在第9章引用的那些），今天依然可以听到。我们引一段最近的颇有感情色彩的评论（Carlo Rovelli, Int.J.Mod.Phys.，D12（2003）：1509-1528）：

我认为弦理论是一个精彩的理论。我对能构筑起这样一个理论的人深感敬佩。然而，一个理论尽管可能令人敬畏，但它在物理上仍然可能是错的。科学史上有许多美妙的思想最终还是错了。我们不能让炫目的数学模糊了双眼。不论弦理论家们无比的才情，激进的革命，还是动人的宣传，这么多年过去了，也没给我们带来什么物理。所有关键的问题依然存在，理论与现实的联系也越来越遥远。从那个理论

导出的所有物理学预言都跟实验相矛盾。将超弦理论看作成功的量子引力理论的老观念我想不再站得住脚了。今天，太多的理论家去拨弄那弦，实在是很大的冒险，无数的心力、一代人的智慧，也许都将浪费在一个美丽虚幻的梦想。

同样的内容，弦学家自己也经常说，不过口气当然不同；似乎弦理论家得意的地方，也是遭反对者批评最多的地方——这大概也是弱势的其他理论家（如圈引力的小圈子中的人）感觉不平衡的地方。我们不知道这样的话还要说到什么时候。

重印前，花几天工夫“匆匆地”重读了一遍，发现了不少或明或暗的错误、遗漏的字句以及排印的失误，尽管多数不会妨碍对主要意思的理解，但总是犯了错误！还有些模糊的、读来费解的地方，严格说来也是错的；对翻译来说，模糊与错误几乎就是同义词。不过这些问题，往往改几个字就能表达清楚。

借改错的机会也改了很多表达方式。例如，删除多余的虚字和过分的形容词，抹去个人发挥的色彩，改变词句的声韵和语调，调整句子的结构和顺序……总之是为了读起来更流畅、更简洁、更好听。老话说，校书如扫落叶，旋扫旋生，永远没有扫干净的时候，既说了事实，也不妨堂皇地拿来做“错误在所难免”的借口。

译者

2004年7月14日

[1]原书的注释在书后。为了方便，我们把简单的说明（如出处）直接当脚注，而把补充性的注释分别放在每一章的后面。有注的地方在右上角以不加圈的数字标记。——译者

[2]前面说过盖尔曼给“夸克”取的名字来自乔伊斯（J.A.Joyce）的《芬尼根守夜者》。在《夸克与美洲豹》（第13章）（《第一推动丛书》第二辑）里，盖尔曼回忆了他是如何想到这个名字的。——译者

[3]希腊神话：Phrygia王Gordius将他的马车献给朱比特（Jupiter），但把轭打了死结，没人解得开。后来亚历山大（Alexander）用剑把结打开了，成了东方的统治者。——译者

[4]爱因斯坦1942年给朋友的信，引自Tony Hey和Patrick Walters的《爱因斯坦的镜子》（Einstein’s Mirror, Cambridge, Eng.：Cambridge University Press，1997）。

[5]例如，P.C.W.Davis, J.R.Brown根据BBC广播节目编辑的一本关于超弦的科普读物就叫Superstrings：A Theory of Everything？（中译本《超弦：一个包罗万象的理论？》廖力、章人杰译，中国对外翻译出版公司，1994年）——译者

[6]Steven Weinberg, Dreams of a Final Theory（New York：Pantheon，1992），p.52。（中译本《终极理论之梦》也收在《第一推动丛书》第三辑中。——译者）

[7]1998年5月11日对E·惠藤的访问。（E·惠藤因为对超弦理论的贡献于1990年获菲尔兹数学奖。——译者）

[8]Van de Graaf（1901～1967）发明了一种静电发生器，办法是在一个绝缘空心金属球上聚集电荷。1931年他在鲁林斯顿建造了第一台发生器。后来他在麻省理工学院将高压发生器发展为粒子加速器。——译者

[9]据作者说，乔治（George）和格蕾茜（Gracie）的名字来自美国著名的George Burns-Gracie Allen喜剧组合，而前面的斯里姆（Slim）和吉姆（Jim）兄弟则没有什么典故。——译者

[10]读者一定要记住，这是许多相对论读物关于“我们看到的”洛伦兹收缩的“传统”错误说法，作者在后面的注释中已经作了补充说明。更具体的讨论可以参考本丛书《时间、空间和万物》的有关章节。——译者

[11]Isaac Newton, Sir Isacc Newton’s Mathematical Principle of Natural Philosophy and His System of the World, trans.A.Motte and Florian Cajori（Berkeley：University of Chi-cago Press，1962），Vol.I, p.634.

[12]引自Albrecht Fölsing, Albert Einstein（New York：Viking，1997），p.315。

[13]John Stachel，“Einstein and the Rigidly Rotating Disk”（爱因斯坦与刚性转盘），in General Relativity and Gravitation（《广义相对论与引力》），ed.A.Held（New York：Plenum，1980），p.1。

[14]引自Fölsing, Albert Einstein, p.189。

[15]1998年1月27日惠勒的谈话。

[16]Robert P.Crease and Charles C.Mann, The Second Creation（New Brunswick, N.J.：Rutgers University Press，1996），p.39.

[17]不过，更令人惊讶的是，最近关于宇宙膨胀速率的详细研究表明，宇宙可能真的藏着一个宇宙学常数，虽然很小，却不是零。——译者

[18]Richard Feynman, The Character of Physical Law（Cambridge, Mass：MIT Press，1965），p.129.

[19]这个比喻源自《圣经·旧约·传道书（10:1）》，原话说，“死苍蝇使做香的油膏发出臭气；这样，一点愚昧也能败坏智慧和尊荣。”——译者

[20]普朗克常数为$1.05\times10^{-27}$克·厘米$^2$/秒或$1.05\times10^{-34}$焦耳·秒。

[21]“量子酒吧”原文是H-Bar，是很巧妙的双关语。Bar既是酒吧，也有“棍”的意思，普朗克常数（h/2π，h上加根棍）在英文里就读h-bar。——译者h

[22]Timothy Ferris, Coming of Age in the Milky Way（New York：Anchor，1989），p.286.

[23]霍金1997年6月21日在阿姆斯特丹“引力黑洞和弦理论”学术会议上的演讲。

[24]Richard Feynman, QED：The Strange Theory of Light and Matter（Princeton：Prin-ceton University Press，1988）.（费曼这本小书的中译本是“商务新知译丛”里的一种，《QED：光与物质的奇异理论》，张钟静译，商务印书馆，1996。——译者）

[25]Stephan Hawking, A Brief History of Time（New York：Bantan Books，1988）.p.175.（霍金的《时间简史》也是我们这套《第一推动丛书》中的一本。——译者）

[26]费曼的话引自Timothy Ferris, The Whole Shebang（New York：Simon＆Schuster，1997），p.97。

[27]我们记得，夸克有一种性质叫“色”，强力的量子力学理论在很大程度上就是描写那些“色”的作用，所以也叫“量子色动力学”，这个名字好像跟“夸克”一样，也是盖尔曼取的。——译者

[28]专业的读者会发现，这一章讲的只是微扰的弦理论；非微扰的理论在第12、第13章讨论。

[29]1997年12月23日J·施瓦兹的谈话。

[30]米谷民秋（Tamiaki Yoneya）以及Korkut Bardakci和Martin Halpern也独立提出这类似的观点。瑞典物理学家Lars Brink对早期弦理论的发展也有过重要贡献。

[31]1997年12月23日，J·施瓦兹的谈话。

[32]M·格林1997年12月20日的谈话。

[33]俄罗斯洋娃娃是嵌套在一起的一组洋娃娃，打开一个，里面还套着更小的。据说这是从日本本州传到俄罗斯的。——译者

[34]美国电视节目The Price Is Right，据作者介绍，这是一档“猜价格”的游戏节目：参与者通过不同的游戏来猜一些商品的价格。——译者

[35]爱因斯坦的话引自R.Clerk, Einstein：The Life and Times（New York：Avon Books，1984），p.287。

[36]Gulliver，斯威夫特（Jonathan Swift，1667—1745）的名著《格列弗游记》（Gulliver's Travels）的主人公。——译者

[37]更准确地说，1/2-自旋的意思是，电子自旋的角动量是h/2。

[38]巴赫（Johann Sebastian Bach，1685—1750）在去世前创作的《赋格艺术》（The Art of Fugue）约包括20首（不同版本数目不同）赋格和卡农（canon），在形式上极尽变化，是高等对位的楷模。最后一首没有完成，曾在艺术史上留下些难题；而从形式看，那对称也就不够完美。——译者

[39]变形虫（音译是“阿米巴”（amoeba））是一种原生动物，最大的不过600微米（长），一般的（如最常见的痢疾内变形虫）只有10微米左右。——译者

[40]Amelia Earhart（1897—1937）是单独飞越大西洋的第一个女飞行员（1932），后来与F.J.Noenan环球飞行时，在太平洋上神秘失踪。——译者

[41]Edward Witten, Lecture at the Heinz Pagels Memorial Lecture Series, Aspen,

Colorado，1997.

[42]关于这一点的更深入的讨论和相关思想，见Steven Weinberg, Dreams of a Final Theory（温伯格，《终极理论之梦》）。

[43]Edwin Abbott, Flatland（Princeton：Princeton University Press，1991）.

[44]爱因斯坦致卡鲁扎的信引自A.Pais, Subtle is the Lord：The Science and the Life of Albert Einstein（Oxford：Oxford University Press，1982），p.330。

[45]爱因斯坦致卡鲁扎的信引自D.Freedman and P.van Nieuwenhuizen，“The Hidden Dimensions of Spactime”，Scientific American 252（1985），62。

[46]爱因斯坦致卡鲁扎的信引自D.Freedman and P.van Nieuwenhuizen，“The Hidden Dimensions of Spactime”，Scientific American 252（1985），62。

[47]爱因斯坦说的科学院即普鲁士科学院，卡鲁扎的文章发表在普鲁士科学院报告（Sitzungsber.d.Preuss.Akad.d.Wiss），1921，p.966。——译者

[48]本图蒙即地密那大学Andrew Hanson同意，用mathematica三维图像软件包绘制。

[49]据作者说，在一个老的美国电视节目中，竞争者回答了所有问题就能赢得64000美元的奖金。——译者

[50]Edward Witten，“Reflections on the Fate of Spacetime”Physics Today, April 1996，p.24.

[51]1998年5月11日惠藤的谈话。

[52]Sheldon Glashow and Paul Ginsparg，“Desperatedly Seeking Superstring？”Physics Today, May，1986，p.7.

[53]Sheldon Glashow的话见The Superworld I, ed, A.Zichichi（New York：Plenum，1990），p.250。

[54]Sheldon Glashow, Interactions（New York：Warner Books，1988），p.335.

[55]费曼的话见Superstrings：A Theory of Everything？ed.Paul Davies and Julian Brown（Cambridge, Eng：Cambridge University Press，1988）。

[56]Howard Georgi的话见The New Physics, ed.Paul Davies（Cambridge：Cambridge University Press，1989），p.446。

[57]1998年3月4日惠藤的谈话。

[58]1998年1月12日瓦法的谈话。

[59]盖尔曼的话引自Robert P.Crease and Charles C.Mann, The Second Creation（New Brunswick, N.J.：Rutgers University Press，1996），p.414。

[60]1997年12月28日格拉肖的谈话。

[61]1997年12月28日格拉肖的谈话。

[62]David Gross，“Superstring and Unification”，in Proceedings of the XXIV Internati-onal Conference on High Energy Physics, ed.R.Kotthaus and J.Kühn（Berlin：SpringerVerlag，1988），p.329.

[63]1997年12月23日施瓦兹的谈话。

[64]1998年3月4日惠藤的谈话。

[65]这一节和后面几节的有些思想不太好懂，如果你跟不上我们解说的思路——特别是，有时是孤零零的一个问题，请你不要泄气。

[66]遗憾的是，2001年9月11日，世界贸易中心大厦在遭恐怖袭击后倒塌了。——译者

[67]K.C.Cole, New York Times Magazine, October 18，1987，p.20.

[68]爱因斯坦的话引自John D.Barrow, Theories of Everything（New York：Fawcett-Columbine，1992），p.13。

[69]跳过第6章“准确的答案”一节的读者，回头去看看那一节的开头应该是有好处的。

[70]1998年5月11日惠藤的谈话。

[71]1996年6月21日，霍金在阿姆斯特丹引力、黑洞和弦学术会议上的演讲。

[72]1997年12月29日格拉肖的谈话。

[73]Laplace, Philosophical Essay on Probabilities, trans.Andrew I.Dale（New York：Springer-Verlag，1995）.

[74]Stephan Hawking, in Hawking and Roger Penrose, The Nature of Space and Time（Princeton：Princeton University Press，1995），p.41（中译本即《第一推动丛书》里的《时空本性》——译者）。

[75]霍金1997年6月21日在阿姆斯特丹引力、黑洞和弦学术会议上的演讲。

[76]1997年12月29日斯特罗明戈的谈话。

[77]1998年1月12日瓦法的谈话。

[78]霍金1997年6月21日在阿姆斯特丹引力、黑洞和弦学术会议上的演讲。

[79]瓦法1998年1月12日的谈话。

[80]1998年5月19日维尼齐亚诺的谈话。

[81]斯莫林的观点见他的书The Life of the Comos（New York：Oxford University Press，1977）。

[82]1998年3月4日惠藤的谈话。

[83]Sir Isaac Newton's Mathematical Principles of Natural Philosophy and His system of the World, trans.Motte and Calori（Berkeley：University of California Press，1962，vol.I, p.6.）

[84]1998年1月12日瓦法的谈话。

[85]1998年5月11日惠藤的谈话。

[86]引自Banesh Hoffman, Helen Dukas, Albert Einstein, Creator and Rebel（New York：Viking，1972），p.18.

[87]Martin J.Klein，“Einstein：The Life and Times, by R.W.Clark”（book review）Science 174，pp.1315～1316.

[88]Jacob Bronowski, The Ascent of Man（Bostion：Little, Brown，1973），p.20.

# 宇宙的结构

# 总序

科学，特别是自然科学，最重要的目标之一，就是追寻科学本身的原动力，或曰追寻其第一推动。同时，科学的这种追求精神本身，又成为社会发展和人类进步的一种最基本的推动。

科学总是寻求发现和了解客观世界的新现象，研究和掌握新规律，总是在不懈地追求真理。科学是认真的、严谨的、实事求是的，同时，科学又是创造的。科学的最基本态度之一就是疑问，科学的最基本精神之一就是批判。

的确，科学活动，特别是自然科学活动，比较起其他的人类活动来，其最基本特征就是不断进步。哪怕在其他方面倒退的时候，科学却总是进步着，即使是缓慢而艰难地进步，这表明，自然科学活动中包含着人类的最进步因素。

正是在这个意义上，科学堪称为人类进步的“第一推动”。

科学教育，特别是自然科学的教育，是提高人们素质的重要因素，是现代教育的一个核心。科学教育不仅使人获得生活和工作所需的知识和技能，更重要的是使人获得科学思想、科学精神、科学态度以及科学方法的熏陶和培养，使人获得非生物本能的智慧，获得非与生俱来的灵魂。可以这样说，没有科学的“教育”，只是培养信仰，而不是教育。没有受过科学教育的人，只能称为受过训练，而非受过教育。

正是在这个意义上，科学堪称为使人进化为现代人的“第一推

动”。

近百年来，无数仁人志士意识到，强国富民再造中国离不开科学技术，他们为摆脱愚昧与无知做了艰苦卓绝的奋斗。中国的科学先贤们代代相传，不遗余力地为中国的进步献身于科学启蒙运动，以图完成国人的强国梦。然而应该说，这个目标远未达到。今日的中国需要新的科学启蒙，需要现代科学教育。只有全社会的人具备较高的科学素质，以科学的精神和思想、科学的态度和方法作为探讨和解决各类问题的共同基础和出发点，社会才能更好地向前发展和进步。因此，中国的进步离不开科学，是毋庸置疑的。

正是在这个意义上，似乎可以说，科学已被公认是中国进步所必不可少的推动。

然而，这并不意味着，科学的精神也同样地被公认和接受。虽然，科学已渗透到社会的各个领域和层面，科学的价值和地位也更高了，但是毋庸讳言，在一定的范围内，或某些特定时候，人们只是承认“科学是有用的”，只停留在对科学所带来的后果的接受和承认，而不是对科学的原动力、科学的精神的接受和承认。此种现象的存在也是不能忽视的。

科学的精神之一，是它自身就是自身的“第一推动”。也就是说，科学活动在原则上是不隶属于服务于神学的，不隶属于服务于儒学的，科学活动在原则上也不隶属于服务于任何哲学。科学是超越宗教差别的，超越民族差别的，超越党派差别的，超越文化的地域的差别的，科学是普适的、独立的，它自身就是自身的主宰。

湖南科学技术出版社精选了一批关于科学思想和科学精神的世界名著，请有关学者译成中文出版，其目的就是为了传播科学的精神，科学的思想，特别是自然科学的精神和思想，从而起到倡导科学精神，推动科技发展，对全民进行新的科学启蒙和科学教育的作用，为

中国的进步做一点推动。丛书定名为《第一推动》，当然并非说其中每一册都是第一推动，但是可以肯定，蕴含在每一册中的科学的内容、观点、思想和精神，都会使你或多或少地更接近第一推动，或多或少地发现，自身如何成为自身的主宰。

《第一推动》丛书编委会

# 第1部分 实在性之舞台

## 第1章 通往实在性之路

空间、时间以及事物为什么是那个样子

我的父亲并不介意别人碰他那布满灰尘的旧书架上的任何一本书。但从小到大，我从未见过任何人从中取下一本。那些书多半是大部头——涉及方方面面的文明史、成套的西方文学巨著以及大量我已记不起来的其他书籍——它们看起来已经和因为数十年牢固的支撑工作而微微弯曲的架子融为一体了。但是在书架的最上一层，有一本薄薄的小书总能引起我的注意，因为它看起来如此不合时宜，就像大人国中的格列佛[1]一样。现在回想起来，很奇怪我那时候竟然没有早点读一读那些书。或许是因为年头太久了，使得那些书不像是用来读的，倒像是祖上流传下来的传家宝一样，令人不敢碰触。不过，这种敬畏之心还是敌不过十来岁孩子不安分的天性。我最终拿起了那本书，拂去表面的灰尘，翻开了第一页。那最开始的几行文字，即使退一步说，也令人非常吃惊。

“只有一个真正的哲学问题，那就是自杀”，这本书就是这样开始的。我有点退缩了。“这个世界是三维空间吗？意识有9种还是12种方式？”这本书提出诸如此类的问题，并解释说这些问题是人类勇敢天性作用下的一部分，但是只有当那个真正的问题解决时这些问题才值得讨论。这本书就是阿尔及利亚哲学家诺贝尔奖得主阿尔伯特·加缪所作的《西绪福斯神话》。后来的一段时间内，随着对他所说的话的逐

渐理解，我心中感到的那丝冰冷才慢慢融化。当然，我觉得你可以一直思考和分析这些问题，但是真正的问题却在于你所有的思考和分析是否使你确信生命值得存在。那才是所有问题的症结所在，其他事情都只是细节而已。

虽然我只是偶然读到加缪的书，但他的话给我留下了极为深刻的印象，时常萦绕耳旁，这远非我读过的其他书所能比拟的。我一次又一次地想：我所遇到的、听说过的或者在电视上看到的各种各样的人会怎样回答这一根本性的问题呢。回想起来，尽管他的第2个论断——关于科学进步所起的作用——对我来说，更具有特殊意义的挑战性。加缪认为理解宇宙结构确实有价值，但是据我所知，他并不认为这种理解可以改变我们对生命价值的评价。现在看来，十几岁的我读有关存在主义哲学的书就像巴特·辛普森[2]读浪漫诗歌一样，但是即便如此，加缪的论断仍然带给了我相当大的震撼。对于这个具有抱负的唯物论者而言，对生命作出明智的评价需要对生命舞台——宇宙有全面的理解。我忍不住想，如果人类居住在深埋在地下的岩石洞穴里，我们就不会发现地球的表面、明媚的阳光、海洋的微风以及远离我们的星球；假设人类沿着一个不同于现在的方向进化：如果人类不能获得除触觉以外的其他感觉，那么我们所知道的一切事物将来自于对周围环境的触觉印象；如果人类大脑发育停止在儿童早期，那么我们的情商和分析能力将不会超过一个5岁的孩子——简而言之，如果我们的经历仅仅是一些对实在性的琐碎描绘——那么我们对生命的评价将大打折扣。当我们最终寻找到通往地球表面的路，当我们具有了视觉、听觉、嗅觉和味觉时，当我们的大脑能够发育到正常水平时，我们对生命和宇宙的看法必然发生根本性的变化。这样看来，我们现有的对实在性的理解将会对所有哲学问题的基础产生非同小可的影响。

会产生什么影响呢？或许你会这样问。任何一位清醒冷静的思考者都会得出这样的结论：虽然我们不可能理解宇宙的每一样东西——关于物质运转和生命功能的方方面面——但是我们仍然可以在大自然

的画布上按照自己的意愿添上粗糙的几笔。确实正如加缪所暗示的那样，物理学上的进步，比如对空间维度数目的理解；或者神经心理学的进步，比如对大脑的所有组织结构的理解；或者，就此而言，其他大量的科学进步都可以说是构成了重要的细节，但它们对我们理解生命和实在性的影响却微乎其微。的确，实在性是我们对世界的认识；实在性是通过我们的经验而展现在我们眼前的。

在某种程度上，我们当中的许多人对实在性都会有上述的看法，只是未明确表述出来而已。我发现自己在日常生活中常以这种方式思考；这样我们就很容易被自然的表象所迷惑。但是，从我第一次读到加缪的书以来的几十年间，我发现现代科学给我们上了与众不同的一课：20世纪以来的科学研究告诉我们，人类的经验往往会对我们理解实在性的本质起误导作用。隐藏在日常生活背后的是一个我们几乎没有多少了解的世界。神秘现象的追随者、占星术的信徒以及那些宣扬超自然的宗教主义的人们，尽管其观点各不相同，但都得到了类似的结论。但那并不是我要说的，我要说的是天才的创造者和孜孜不倦的科学家们的工作，如同剥洋葱一样，揭开了宇宙一层层的面纱，探索了一个又一个的难题，向我们展示了一个完全不同于平常人们所认识的宇宙，一个令人惊奇、兴奋和优雅的宇宙。

但是这些科学进展也只是细节而已。自然科学的重大突破已经驱使并且有力量继续驱使我们的宇宙观发生戏剧性的变化。现在，我仍然像几十年前一样确信，加缪把生命的价值作为基本问题讨论是正确的，但现代自然科学让我相信，通过日常经验来评价生命就像通过一个空可乐瓶来凝视凡·高一样。作为先锋的现代科学，向我们的基本感知发起了一轮又一轮的攻击，使我们对我们存于其间的这个世界产生了很多概念性的迷惑。因此，即便加缪把物理问题分离出来并置于从属地位，我仍然相信科学才是最根本的。在我看来，科学的实在性为反驳加缪的观点提供了舞台和一些启示。评价存在性问题而忽略了现代科学的洞察力就像在黑暗中与一名不知名的对手摔跤。加强对科学

实在性的真正本质的理解，将有利于重塑我们的人生观和宇宙观。

这本书的核心在于阐释对实在性的理解起最为关键作用的一些修正，主要集中在那些长期以来影响人类时空观念的工作上。从亚里士多德到爱因斯坦，从罗盘到哈勃望远镜，从金字塔到山顶上的天文台，自从思想产生时，时空就为我们的思想确立了框架。随着现代科学的到来，它们的重要性已大大提高。近3个世纪以来，物理学的发展表明，时空观念已被看做最令人困惑且最引人注目的问题，但同时也是对宇宙进行科学分析的基础；时空观念已列于古老的科学概念之上，而这些古老的科学概念被边缘研究改建得更加奇妙。

对于艾萨克·牛顿而言，空间和时间只不过是一个永恒不变的、普适的宇宙舞台，以便宇宙中的事件能够在此一幕幕地上演。对于同时代可与牛顿匹敌的戈特弗莱德·威廉·范·莱布尼茨而言，“空间”和“时间”只是与物体在哪里和事件何时发生有关系的词汇。换句话说，空间和时间对他们而言并不能代表什么。但对于阿尔伯特·爱因斯坦而言，空间和时间是隐藏在实在性下面的原始材料。通过相对论，爱因斯坦震撼了我们的时空观，向我们展示了它们在宇宙演化中所起的重要作用。从那时起，空间和时间就成为物理学界最耀眼的明珠。对我们而言，时间和空间是既熟悉又神秘的；彻底地理解时间和空间已成为最令物理学家胆怯的挑战，而它同时也是受欢迎的猎物。

本书所讲述的物理学进展主要是一些关于时空结构的不同理论。其中的一些理论挑战了人类长久以来——即便没有几千年也有几百年——的时空观。另外一些理论则将试图寻找我们对时空理论的理解和日常生活经验之间的联系。而其他理论将质疑普通观念限定范围内所不能解释的一些问题。

我们在本书中将尽可能少地涉及哲学问题（这里当然不是指关于自杀和生命的意义的那些哲学）。但在探寻科学解释空间和时间之奥妙的过程中，我们完全是不受任何限制的。从宇宙最小的微粒和最早

的时刻到其所能达到的最远的边界和最遥远的未来，我们将在熟悉和宽广的环境中探索时间和空间，不停歇地去追求其真正性质。科学家们对空间和时间的探索还在继续，此刻我们还无法作出最终的评价。我们将会介绍一系列的进展——有些非常奇怪，有些令人非常满意，有些已被实验证实，有些还只是空想——它们将向我们展示人类对宇宙结构的思考究竟到了何种地步，人类的指尖对实在性真正纹理的触摸已到了何种深度。

## 经典意义上的实在性

关于现代科学从何时开始，历史学家们众说纷纭，尚无定论。但毫无疑问的是，从伽利略、笛卡儿、牛顿等人开始创造他们的学说时起，现代科学已经走上了正轨。在那个时代，新的科学意识体系正在稳步地建立起来，地球上和天文学上实验数据的规律性使人们越来越清楚地看到，宇宙的过去和未来是有规律可循的，通过精密的推理和数学分析我们可以找到这些规律。富有现代科学思想的早期先驱们指出：回顾走过的科学之路，宇宙中发生的事件不仅可以解释，而且也可以预测。科学所具有的预言未来方方面面的力量——持续而定量的——早已得到了证实。

早期的科学研究主要集中在我们日常生活中可以看到或体验的各种事物。伽利略从斜塔上抛落重物（大约是人所共知的传奇），或者观察沿斜面滚落的小球的运动；牛顿研究树上落下来的苹果（又是一段传奇）和月球轨道。这些研究的目的在于使新生的科学研究与大自然和谐一致。当然，物理学中的实在性是各种体验的来源，但更富挑战的是聆听自然的和谐之声并寻找隐藏在其背后的原因。许多著名学者和无名英雄都为早期科学的飞速发展作出了巨大贡献，但最后只有牛顿成了舞台上的明星。通过对数学方程的运用，牛顿将地球和天空中的各种已知运动现象综合了起来。就这样，今日所谓的经典物理学诞生了。

在牛顿完成其工作之后的几十年间，他的方程被发展出了详尽精密的数学结构，大大丰富了原始理论，扩展了其实际用途。经典物理学逐渐成为一种深奥精妙而又成熟的科学体系。但是照亮科学之路的却是牛顿富有创造性的洞察力。即使今天，300多年过去了，我们可以发现牛顿方程依然出现在世界各地的初级物理学课程中；在NASA（美国国家航空和宇宙航行局）的飞行计划里依然用牛顿方程来计算太空船的运行轨迹；在前沿研究的复杂计算中也常常有牛顿方程的一席之地。在一个单独的理论体系下，牛顿带来了丰富的物理学现象。

但在总结他的运动学定律时，牛顿遇到了一个棘手的问题，而这个问题对于我们所要讲述的故事也很重要（第2章）。每个人都知道物体可以运动，但是这些运动发生在哪里呢？空间，也许大家都会回答。但是，牛顿却会问，空间又是什么呢？空间是一个真正的物理实体？还是人们根据对宇宙的理解而得出的一个抽象概念呢？牛顿意识到这个关键问题必须得以解决，因为如果没有对空间和时间的正确理解，他的公式将变得毫无意义。理解需要来龙去脉，思考需要正确的方向。

因此，在他的《数学原理》一书中，牛顿用简明的语言阐述了空间和时间的概念，他认为空间和时间是绝对的、不可改变的实体，这就为宇宙提供了一个固定而不可改变的舞台。根据牛顿的理论，空间和时间为宇宙提供了一个不可见的框架从而形成了宇宙结构。

即使在当时，也并不是每个人都同意牛顿的说法。有些学者就指出把理论建立在你摸不到又看不着也无法影响的事物上是没有意义的。但是牛顿的解释和牛顿方程惊人的预言能力使这样的观点销声匿迹。在之后的200年里，牛顿关于空间和时间的绝对性观点成为铁律。

相对论意义上的实在性

经典的牛顿世界观之所以令人心悦诚服，并不仅仅是因为它能以惊人的精确度描述自然现象，更是由于这种对大自然的描述的细节之处——数学——是与经验紧密相连的。你用力推一个物体，它就会加速。你掷出球时花的力气越大，球撞墙时所发生的形变也就越大。当你挤压某个物体时，你也会感觉到那个物体在挤压你。一个物体的质量越大，它所具有的重力也就越大。所有这些都是自然世界的最基本性质。当你学习牛顿的力学体系时，你会发现所有的这一切都可由牛顿的方程清晰直观地表示出来。不同于用水晶球占卜那一套完全无法了解的骗人伎俩，任何一个只受过很少数学训练的人都可以掌握牛顿定律。经典力学为人类的直觉提供了坚实的基础。

引力很早就被牛顿纳入其方程之中。但直到19世纪60年代，电力和磁力才由苏格兰物理学家詹姆斯·克拉克·麦克斯韦添加到经典物理体系中。麦克斯韦需要使用新的方程来描述电力和磁力，而他所用到的数学知识需要更高层次的训练才能完全掌握。由麦克斯韦引入经典物理体系中的新的方程在描述电磁现象上恰如牛顿方程在描述运动上那样成功。到了19世纪末，宇宙的奥秘显然已不是人类智力的对手了。

事实上，随着电与磁的成功统一，科学家们逐渐产生了一种认识：理论物理即将完善。有人提出，物理学正在飞速地发展为一门完善的学科，它的定律不久后就会被雕刻在石碑上。1894年，著名的实验物理学家阿尔伯特·迈克尔逊评论道，“大多数重要的基本理论已牢固地建立起来”，他引用了一位著名的科学家的话——大多数人认为这句话是英国物理学家罗德·开尔文说的——物理学界剩下的工作只是确定小数点后的数字之类的问题。[1]1900年，开尔文曾指出物理学界上空盘旋着两朵乌云，一个与光的运动性质有关，另一个则是物体被加热时的辐射问题，[2]但在当时，大家都觉得这些也仅是一些细节问题，它们很快就会被解决。

在随后的10年间，一切都改变了。虽然正如人们所预料的那样，

开尔文提出的这两个问题很快被解决了，但是它们的解决却带来了更多的故事。每个问题的解决都导致了一场革命，都需要改写基本自然定律。空间、时间和实在性——几百年来它们不仅有效运转，而且精确地表达了我们对世界的直觉——将不得不被丢弃了。

1905～1915年，阿尔伯特·爱因斯坦完成了狭义相对论和广义相对论，掀起了一场解决开尔文第一朵“乌云”的革命（第3章）。在电、磁和光的运动的谜团中挣扎时，爱因斯坦意识到，经典物理学的基石——牛顿的空间和时间概念出现了状况。通过1905年春季几个星期的努力，爱因斯坦提出：空间和时间并不像牛顿认为的那样具有独立且绝对的存在性；两者实际上以一种与日常经验相反的形式相互联系。10年之后，爱因斯坦重写了引力定律，为牛顿定律的棺木敲上了最后一颗钉。这次，爱因斯坦指出空间和时间不仅是一个统一整体的一部分，而且通过自身的蜷曲参与了宇宙演化。空间和时间远非如牛顿所想象的那样具有稳固且不可改变的结构；在爱因斯坦的理论中，它们富有弹性并且可以不断变化。

狭义与广义相对论是人类最宝贵的成就，爱因斯坦正是利用它超越了牛顿体系中实在性的概念。即使牛顿的经典物理学在数学上看起来与我们所能感知的物理世界相符合，但它所描述的实在性并不是我们世界的实在性。我们生活于其中的乃是一个具有相对论意义上的实在性的世界。但是，由于经典物理与相对论物理的实在性之间的偏离只有在极端的情况下（如在速度和引力非常大时）才非常明显，因此在大多数情况下，牛顿理论作为一种近似，仍然具有一定的精确性及有效性。但功用性和实在性是完全不同的标准。我们将会看到，我们习以为常的空间和时间的性质只不过是错误的牛顿式臆想而已。

## 量子世界中的实在性

罗德·开尔文提出的第二种反常为我们带来了量子革命，它是现代人类在对宇宙的认知上不得不承受的一场剧变。当这场剧变烟消云

散、经典物理学的饰面被烧焦后，浮现出了新鲜的量子实在性的理论框架。

经典物理学的一个关键特性在于，如果你知道某一时刻所有物体的位置和速度，那么根据牛顿方程和麦克斯韦添加的新方程，你就可以推算出任意其他时刻（包括过去和未来）所有物体的位置和速度。毋庸置疑，在经典物理学体系下，过去和未来都可以与现在准确地联系起来。而狭义相对论和广义相对论也有此一说。虽然在相对论框架下，过去和现在的概念比我们所熟知的经典物理中的过去和现在（第3章和第5章）的概念要微妙一些，但只要有了相对论方程以及现在的现在这一时刻的所有物理条件，我们还是能够推导出有关过去和未来的一切。

而到了20世纪30年代，物理学家们被迫引进一种全新的理论体系——量子力学。令人意想不到的是，人们逐渐发现只有依靠量子定律才可以解开大量谜团，并为其时刚刚从原子和亚原子层次测得的各种数据找到合理的解释。但是根据量子定律，即使你对现在的事物的状态做出最完善的测量，你最多也只能预言物体在未来或过去某个时刻的运动路径的概率。根据量子力学，宇宙并不能从现在完全推演出，我们所能得到的只是概率。

虽然对于如何解释这些进展仍存在着争议，但大多数科学家都认同概率的概念与量子意义上的实在性是密不可分的。人类直觉及其在经典物理中的表现形式会勾画出一幅有关实在性的图像，在这幅图像中，物体总会明确地朝着这个或那个方向运动；但量子力学则不然，它所描绘的只是物体徘徊在各种运动状态之间的实在性。只有当某种观测事件强迫物体放弃量子概率时，我们才会得到确定的结果。而尽管我们不能预言这最终的结果，但我们却可以预言物体处于或这或那的运动状态的概率。

坦白地说，这听起来的确有点不可思议。我们并不习惯这种在具

体的观测之前一直保持模糊的实在性。但是量子力学的奇异性并未到此为止。1935年，爱因斯坦和两个年轻同事内森·罗森和鲍里斯·波多斯基写的一篇论文令人又一次震惊于量子力学的奇妙，在这篇论文中，作者们试图发起一次对量子理论的攻击。[3]基于科学发展的螺旋式上升说法，我们可以将爱因斯坦的论文看做是第一篇指出了量子力学——从表面意义上来说——隐含着某种可能性的文章。这种可能性即在此地发生的事情可以瞬时地与彼地发生的事情联系起来而不用考虑距离的问题。爱因斯坦认为这种瞬时联系荒谬可笑，并将这种来自于量子理论的数学结果视作证明量子理论仍有待发展的明证。但是到了20世纪80年代，理论和工程技术的极大发展使得用实验来检验这种假想的量子谬论成为现实，研究者们证实了相距甚远的两个不同位置间发生瞬时联系的可能性。在早期实验条件下令爱因斯坦感到荒谬可笑的事情真的发生了（第4章）。

关于量子力学特性对实在性图像的影响是一个正在进行中的研究课题。许多科学家，包括我自己，将它们看做是关于空间意义与性质的激进的量子升级中的一部分。正常看来，空间间隔意味着物理上的独立性。如果你想要控制足球场那边发生的事情，你不得不去那儿，或者，至少，你不得不送某人或某物（助理教练，可以传播声音的空气分子，或能引起注意的闪光，等等）穿越球场以施加你的影响。如果你什么都不做——如果你保持原地不动——你将不会对球场的那边产生任何影响，因为球场中间的大片空间将阻断任何物理联系。而量子力学，至少在某些情况下，通过展现超越空间的能力对这种观点提出了挑战；大范围的量子关联可以避开空间间隔。两个物体在空间可以相隔很远，但从量子力学的角度考虑，它们似乎就可看做一个整体。而且，由于爱因斯坦所发现的空间和时间之间的紧密联系，量子关联对时间也有影响。我们将在后面的章节中介绍一些巧妙又真正神奇的实验，这些实验最近探测到量子力学中一些令人非常吃惊的时空关联，正如我们所看到的，它们非常有力地挑战了经典的、我们大多数人所持的直觉性观点。

除了上述的那些令人印象深刻的观念，还有一个有关时间的基本特性——它的方向是从过去指向未来的——相对论和量子力学都没法给出解释。关于这一问题，唯一令人信服的进展来自于物理学领域名为宇宙学的研究。

宇宙学里的实在性

睁大你的双眼寻求宇宙的真正本质一直是物理学的最基本目的之一。很难想象有什么经历会比认识到——如我们20世纪做到的那样——我们所感受到的实在性只不过是实在性的一缕而已，更能令人觉得不可思议的了。但物理学的另一个同等重要的目标就是解释我们在实际生活中所感受到的实在性。回顾一下物理学史，似乎这个目标已经达到了，似乎普通的生活感受已经被20世纪前的物理学解决了。从某种程度来看，这是正确的。但即使是那些日常之事，我们也远远没有完全理解。我们日常感受到的很多事情都仍未被解释清楚，其中之一即是现代物理学界最深奥的秘密——伟大的英国物理学家阿瑟·爱丁顿爵士把它称为时间之箭。[4]

我们理所当然地认为万物都有一个时间上的发展方向。鸡蛋一旦打碎就不再完整了；蜡烛一旦融化就不能再重塑起来了；记忆一旦成为过去就不再属于未来；人们一旦年老就不再年轻了。这些不对称性主宰了我们的生活；过去和未来的区别是检验实在性的主要因素。如果过去和未来表现出来的对称性与我们所见证过的左与右，后与前的对称性一样，那么这个世界将变得无法认知。每当鸡蛋打碎时，碎片很快就可连接起来；每当蜡烛融化时，蜡油很快就可重塑；我们会记起很多未来的事情就像我们回忆过去一样；每当人们变得年老时很快就会再变得年轻。显然，这种时间上的对称性并不具有实在性。但是时间的不对称性来自哪里呢？是什么决定了时间所有的最基本的特点呢？

事实上，一些被广泛接受的著名物理定律并未显示出这样的不对

称性（第6章）：时间的每个方向，向前或向后，在定律中都是没有区别的。而这是一个巨大谜团的起源。基础物理学的方程中并无时间方向的区别，这与我们的日常生活经验是不一致的。[5]

令人惊奇的是，尽管我们的注意力集中于熟悉的日常生活，但要想令人信服地解决基础物理与基本体验之间的不相容，我们就得思考最不熟悉的事情——宇宙的起源。这种认识植根于19世纪物理学家路德维格·玻尔兹曼的工作，许多研究者对此进行过详尽的说明，其中最著名的有英国数学家罗杰·彭罗斯。我们将会看到，宇宙开端时的特殊物理环境（在宇宙大爆炸之时或之后的高度有序的环境）可能已经为时间选择了一个方向，正如将一个时钟的发条拧紧，使之处于高度有序的初始状态，然后它就会滴滴答答转起来。因此，从某种意义上讲，鸡蛋破碎——而不是破碎的鸡蛋重新完整——见证的是140亿年前宇宙诞生时即已设定的条件。

日常经验与早期宇宙之间出乎意料的联系使我们认识到为什么事物总是向时间轴的一个方向发展下去而不能反过来进行，但是这并未完全解决时间之箭的神秘性。相反，它把谜团扩大至整个宇宙学——针对宇宙起源与整个宇宙的演化的研究——它促使我们寻找宇宙是否有高度有序的开端，而这正是解释时间之箭的关键所在。

宇宙学是令我们人类感到困惑的最古老学科。这一点都不奇怪。我们是讲故事的人，什么故事能比有关创生的故事更伟大呢？过去的几千年间，世界范围内的宗教和哲学教义对万物——宇宙——的起源提出过种种不同版本的诠释。科学，在其漫长的历史中，也早已染指宇宙学。但直到爱因斯坦的广义相对论出现之后，现代科学宇宙学才真的诞生。

在爱因斯坦发表他的广义相对论之后不久，许多科学家，包括他自己，都曾将广义相对论应用于整个宇宙。十几年后，他们的研究导致了试探性的理论体系即所谓大爆炸理论的诞生，该理论成功地解释

了许多天文学观测现象（第8章）。20世纪60年代中期，支持大爆炸理论的证据进一步增多，通过观测，人们发现空间中存在着近似均匀的微波辐射——虽然肉眼不可见但很容易就能被微波探测器检测到——而这是大爆炸理论预言过的。到20世纪70年代，经过更为细致的研究，人们取得了实质性进展，已经能够确定不同时期热量与温度上的改变与宇宙的基本组成之间的对应关系，大爆炸理论稳固了其在宇宙学理论中的领先地位（第9章）。

尽管其成功毋庸置疑，但大爆炸理论也有非常明显的缺陷。它无法解释太空中为什么会有天文学观测所发现的整体形状，也无法解释为什么微波辐射的温度在整个天空中各向同性——自打微波背景辐射被发现起人们就一直在研究的问题。而且，从我们研究的最初目的来看，大爆炸理论并未提供令人信服的理由来解释宇宙在最初时刻高度有序的原因，而这正是解释时间之箭的关键所在。

20世纪70～80年代，各种各样未被解决的问题引发了一场重大突破，导致了所谓的暴涨宇宙学（第10章）的诞生。暴涨宇宙学修改了大爆炸理论，在宇宙最初时刻插入了一场以令人难以置信的速度急剧膨胀的极短暂的爆炸时期（在该理论中，宇宙在不到百万亿亿亿分之一秒的时间内，其大小增加了百万亿亿亿倍）。年轻的宇宙的疯狂增长填补了大爆炸模型所留下的空隙——从而解释了宇宙形状和微波辐射的均匀性，也暗示了早期宇宙高度有序的原因——于是也就在解释天文学观测和我们所体验到的时间之箭方面取得了实质性进展。

然而，尽管取得了以上成功，但发展了20年的暴涨宇宙论一直隐匿着它令人遗憾的一面。与其所修正的标准大爆炸理论一样，暴涨宇宙论需要借助于爱因斯坦所发现的广义相对论方程。虽然大量的研究文章证明了爱因斯坦方程在精确描述大型重量级物体上确实有效，但是长久以来物理学家们都知道有关小物体的精确理论分析——比如所观测宇宙的年龄还不到1秒时——需要使用量子力学。问题是当广义相对论的方程与量子力学的方程混合在一起时，将发生灾难性的结

果。方程完全破产了，这将阻碍我们确定宇宙的诞生过程，以及其在诞生之初时的条件是否能用来解释时间之箭。

将这种情况描述成理论学家的梦魇并非夸张：分析一个实验上无法涉及的重要领域时却找不到数学工具的帮助。因为空间和时间与这一特别的难以企及的领域——宇宙起源——有着千丝万缕的联系，所以要理解空间和时间，我们就得找到这样一种方程——它得能应付早期宇宙高密度、高能量、高温度的极端条件。这绝对是一个具有本质意义的目标，许多物理学家相信，要解决这个问题需要发展所谓的统一理论。

统一理论的实在性

过去的几个世纪里，物理学家们通过说明各种各样千变万化的或者在表面上看来全无联系的现象实际上可归结于单独一组物理定律，来肯定我们对自然界的理解。对于爱因斯坦而言，这样统一的目标——用最少的物理原则来解释最广泛的现象——成了他延续一生的激情。借助于他的两个相对论，爱因斯坦统一了空间、时间和引力。但是这种成功只能起到鼓舞他，令他思考得更加深远的作用。他梦想着能找到一个独立的，能涵盖所有自然定律的理论体系；他将这样的理论体系称为统一理论。虽然时不时地有谣传说爱因斯坦已经发现了统一理论，但是所有的传言都毫无根据；爱因斯坦的梦想还没有实现。

在爱因斯坦一生的后30年中，他集中精力研究统一理论而使他脱离了当时的主流物理学。在许多年轻科学家看来，对于一位像爱因斯坦这样的伟人而言，一意孤行的研究统一理论是走错了方向。但是在爱因斯坦逝世后的几十年里，越来越多的科学家接过爱因斯坦的接力棒，继续他未完成的事业。今天，发展统一理论已成为理论物理学界最重要的问题。

许多年来，物理学家们发现实现统一理论的关键障碍在于20世纪

物理学界的两个重大突破——广义相对论和量子力学——之间的矛盾。虽然这两个体系应用于不同的领域——广义相对论应用于宏观物体如星球和星系，而量子力学应用于诸如分子原子之类的微观物体——但是它们各自都宣称自己具有普适性，可以应用于所有的领域。不管怎样，正如前面所说，一旦将这两个理论结合起来，所得到的方程就会产生毫无意义的答案。比如，当用量子力学结合广义相对论来计算某个过程或与引力有关的事件发生的概率时，答案不是24%、63%、91%之类；相反，所得到的数学结果居然是无限大。但这并不意味着某事发生的可能性如此之大，必胜无疑，以至于你应该把所有的钱都赌进去。大于100%的概率是没有意义的。计算得出无限大的概率只能说明将广义相对论和量子力学结合起来所得到的方程是有问题的。

科学家们意识到相对论和量子力学之间存在矛盾已有半个多世纪了，但在相当长的时间内，很少有人感到有必要去解决这个问题。相反，大多数研究者用广义相对论来分析大而笨重的物体，而用量子力学来分析小而轻的物体，他们小心翼翼地在每个理论的安全范围内使用它们，防止了相互矛盾的产生。许多年来，这种绥靖政策使得这两种理论分别在各自的领域取得了相当大的成就，但这并不能带来永久的和平。

很少的一些领域——既具有大质量又具有小尺度的极端物理条件下——即那些需要广义相对论和量子力学同时有效的领域彻底沦为军事禁区。举两个最为人所熟悉的例子：在黑洞中心，一个完整的星球由于自身重力可被压缩成一个微小的点；大爆炸理论假想整个可观测宇宙被压缩成比一个原子还要小很多的核。没有广义相对论和量子力学的成功结合，塌缩星球和宇宙起源将永远是未解之谜。但许多科学家宁愿把这些领域搁置一边，或至少把其他易解决的问题解决之后再来考虑这个问题。

但是有些科学家却无法等待。公认的物理定律之间的矛盾意味着

掌握深层次真理的失败，这已经足以使这些科学家无法安心了。尽管路途艰辛，但那些不断探索的科学家们却发现了更加深不见底的水域和更为汹涌的波涛。然而随着时间的慢慢逝去，研究却无多大的进展；一切看起来仍是那么虚无缥缈。尽管如此，那些坚定决心探索这个领域并不断追寻着统一广义相对论和量子力学的科学家们仍然值得嘉奖。科学家们现在正沿着被先驱们照亮的道路继续前进，已经快要实现这两个理论的完美结合。大多数人认为最有竞争实力的理论是“超弦理论”（第12章）。

我们将会看到，超弦理论发轫于为一个老问题给出的新答案：最小的，不可再分的物质是什么？几十年来，传统答案是物质是由粒子组成——电子和夸克——它们被模型化为不可再分，没有大小也没有内部结构的点。传统的理论认为——并有实验证实——这些粒子以各种不同的方式结合起来产生质子、中子和各种各样的原子和分子，进而组成我们平常肉眼所见的各种物体。超弦理论是一个与众不同的理论。它并未否认电子、夸克和实验所发现的其他粒子所起的重要作用，但它声明这些粒子并非是点。相反，根据超弦理论，每种粒子都是由极小的能量丝组成，它们比单个原子核还要小万亿亿倍（远非我们目前所能探测的长度），这种能量丝形成一种形似弦的东西。就像小提琴的弦能按不同的模式振动，而每种振动模式都能产生不同的音调一样，超弦理论中的能量丝也有多种振动模式。这些弦的振动尽管不会产生不同的音调；但根据该理论，它们却会产生不同的粒子性质。按某种模式振动的细小的弦有电子的质量和电荷；据该理论，这就是我们传统意义上的电子。按不同模式的振动着的另一根小小的弦将会是夸克、中子或其他类型的粒子。所有不同类型的粒子都可以在超弦理论中得到统一，因为每种粒子都是由相同基本实体的不同模式的振动产生的。

表面看来，从点演变到极小以至于看起来像点的弦似乎并不具有重大意义。但事实并非如此。从这样微小的开始，超弦理论把广义相

对论和量子力学整合为一个独立、连贯的理论，从而消除了以前尝试统一所得出的无限大的概率。如果那还不够的话，超弦理论已经展现出了将自然界中所有的力和所有的物质都统一于同一理论的必然性。简而言之，超弦理论是爱因斯坦统一理论的重要候选者。

这些伟大的主张，如果正确，将意味着里程碑式的进步。但是超弦理论最激动人心的特征——毫无疑问，这也将使爱因斯坦激动——在于其对我们理解宇宙结构上所产生的重大影响。我们将会看到，根据超弦理论，我们的时空观必须来一次巨变，才有可能在数学上合理地将广义相对论和量子力学结合起来。超弦理论要求有9个空间维度和1个时间维度存在，而不是常识中的三维空间和一维时间。在超弦理论更加复杂的化身所谓的M理论中，统一需要十维空间和一维时间——宇宙在根本上有11个时空维度。由于我们看不见这些额外的维度，超弦理论相当于告诉我们，我们迄今为止所瞥到的实在性只有那么一点点。

当然，没有额外维度的观测证据也可能意味着它们并不存在以及超弦理论是错误的。然而就这样得出结论是十分草率的。在发现超弦理论的几十年前，一些富有前瞻性的科学家，包括爱因斯坦，曾经思考过这个问题，或许空间维度远远不止我们所看到的那么多，额外的维度可能隐藏起来了。研究弦论的科学家们充分精炼了这些思想并发现额外维度很难取得突破性进展，因为它们太小以至于无法用现有的设备看到（第12章），也可能它们很大但是我们探索宇宙的方式无法将其发现（第13章）。每一种猜想都会带来丰富的内涵。通过其弦的振动的影响，微小且有褶皱维度的几何形状在回答大多数基本问题（如为什么我们的宇宙中存在着恒星和行星）上起着关键性的作用。大的额外空间维度所留有的想象余地允许更不可思议的想法：或许存在着其他的，我们至今完全没有察觉到的周边——不是普通空间意义上的周边，而是额外维度上的周边——世界。

虽然只是一个大胆的想法，但额外维度的存在并不只是理论上天

马行空的想法。它也许很快就会得到验证。如果额外维度真的存在，我们就有可能在下一代原子对撞机上得到意想不到的结果，比如说人类将有可能第一次人工合成微观黑洞，或是产生一大批新的、从未被发现的粒子（第13章）。各种各样的奇异结果可能为平常不可见的额外维度提供第一手证据，并使超弦理论离人们长久寻求的统一理论更进一步。

如果超弦理论被证明是正确的，那么我们将不得不接受我们所知道的实在性不过是厚重又纹理丰富的宇宙织物的精美花边而已。尽管据加缪看来，确定空间维度的数目——特别有意义的是，如果真的发现空间维数并不只是3个——所提供的远不只是一些科学上很有趣但却最终无多大意义的细节。但额外维度的发现将使我们认识到，全部的人类经验还不足以帮助我们完全知晓宇宙在基本层面上实质性的东西。这将不容辩驳的证明，我们过去以为通过人类感官即可轻易明了的宇宙特性也不一定就如我们认为的那样。

过去和未来的实在性

随着超弦理论的发展，科学家们非常乐观地认为我们已经建立起了一个任何情况下——无论情况多么极端——都将成立的理论体系；而有了这个理论，我们就可以在未来的某一天，在这个理论方程的帮助下，搞清楚宇宙起源的那一刻究竟发生了些什么。迄今为止，还没有人有办法将该理论合理地应用于大爆炸理论，但是从超弦理论的角度来理解宇宙已经成为当代研究的首选之一。过去的一些年里，有关超弦理论的世界范围内的研究计划已经为我们带来了新颖的宇宙学体系（第13章），使我们可以用天文学观测来验证超弦理论（第14章），并使我们有机会初窥超弦理论在解释时间之箭中应起的作用。

时间之箭，因其在我们日常生活中所扮演的重要角色及其与宇宙起源之间的密切联系，而处在我们所感知的实在性与前沿科学试图追寻的更为精确的实在性之间的交汇地带。如上所说，时间之箭的问题

串起了我们将要讨论的许多科学的进展，有关时间之箭的问题将会在后面的章节中反复讨论，这样才是恰当的。在影响我们生活的许多因素中，时间占有最重要的地位。当我们能熟练掌握超弦理论及其扩充——M理论时，我们对宇宙学的理解将会更加深刻，对时间起源与时间之箭的理解也将更为锐利。如果我们能让想象自由驰骋，我们甚至能预想到我们理解的深度终有一天能让我们自由地探索时空，进而探索那些远非我们现在的能力所能达到的领域（第15章）。

当然，我们可能永远都没办法拥有这种力量。但即使我们从来都没有控制时间和空间的能力，深刻的理解也会使我们有自己的力量。我们对时间和空间本性的理解将是人类智力的证明。我们最终会了解空间和时间——这一悄然矗立着的限定人类感知范围的界碑。

空间和时间的下一个时代

许多年前，当我翻到《西绪福斯神话》的最后一页时，我就对文中所体现出的无上的乐观情绪感到惊奇。一个人被诅咒，当他把一块大石头推上山时，这块石头将会滚下来，于是他又不得不再次将石头推上山，这毕竟是那种注定不会有幸福结局的故事。但是加缪却在他身上找到了真正的希望，西绪福斯追求自由的意志，勇敢地面对不可逾越的障碍，英勇地选择了生存，即使他因被诅咒而不得不在冷漠的宇宙里做一项十分荒谬的工作也不放弃。西绪福斯放弃除即时体验外的一切，不去寻找任何一种更深刻的理解或更深刻的意义；在某种意义上，加缪认为，西绪福斯胜利了。

对于在别人看来只有绝望的事情，加缪却有能力发现其中的希望，我对此感到十分震惊。但是作为一个少年，即使在几十年后的今天依然如此，因为我无法同意加缪的观点——对宇宙的更深刻的理解将不会使生活更有意义。西绪福斯是加缪心目中的英雄，而我心中的英雄却是最伟大的科学家——牛顿、爱因斯坦、尼尔斯·玻尔和理查德·费曼。我读过费曼对于玫瑰的描述——他解释他如何感受到玫瑰的芳

香和美丽，就像其他人做的那样，而他的物理知识大大丰富了人类的体验，因为他也能理解基本的分子、原子和亚原子过程中的奇迹和壮丽——我被他深深地迷住了。我推崇费曼描述的：在所有可能的层面上体验生活、感知宇宙，而不仅仅停留在那些恰好符合人类感知能力的层面上。寻求对宇宙的深层次理解已经成为我活力的源泉。

作为一名物理学教授，长久以来我认识到我高中时在对物理的沉迷中存在一些天真的想法。物理学家们一般不会怀着对宇宙的敬畏和幻想把时间花在思考花儿上。相反，我们把许多时间花在研究爬满黑板的复杂数学公式上。进展是如此的缓慢。即使富有希望的想法也往往毫无所得。那就是科学研究的本质。但是，即使在只有微小进展的时期，我也会发现在猜测和计算上的努力使我同宇宙之间的联系变得更加紧密了。我觉得一个人开始了解宇宙，并非只能通过解释宇宙的神秘之处了解，也可以通过令自己沉浸在研究宇宙的乐趣之中来了解。获得答案很了不起，获得被实验验证的答案就更了不起。但即便最终被证明是错误的答案，也代表着与宇宙的联系加深了一步——这种联系为我们的问题带来了更多的光亮，进而使我们对宇宙了解更近了一步。即使与特定的科学探索有关的巨石依然滚回原地，我们还是会学到一些东西并丰富我们对宇宙的体验。

当然，科学史向我们揭示，人类集体科学探索这块巨石——几个世纪以来各个大陆的无数科学家们所作出的贡献——不会从山上滚下来。不像西绪福斯，我们不用一次又一次从头再来。每一代科学家们从先辈手中接过接力棒，对先人的艰苦工作、洞察力、创造性表示敬意，并把它们再向前推进一点。新的理论和更为精准的测量是科学进步的标志，而这样的进步又建立在以前科学成就的基础上。正是由于这样的原因，我们的任务并非荒唐而无意义的。在把巨石推向山顶时，我们承担的是最精细、最高贵的工作：揭开我们称之为家园的这片地方的奥秘，在我们发现的奇境中畅游，把我们的知识传授给我们的跟随者。

作为一个能直立行走的物种，我们所面临的挑战是艰难的。在过去的300年中，我们取得了巨大的进步，从经典物理学到相对论，再到量子的实在性，直到今天对统一实在性的探索，我们的思想和工具都经历了时间和空间上的巨大跨越，使得我们离这个伪装得如此之巧妙的世界更近了。我们正在逐步揭开宇宙的面纱，一步步接近真理。虽然我们的探索之路还很长，但很多人相信，人类已经快要结束童年时期了。

可以肯定的是，我们对银河系[6]的探索已经酝酿了相当长的一段时间。我们以这样或那样的方式探索这个世界、思考整个宇宙已经有几千年了。大多数时候，我们只简单地探索未知之物，每一次尽管稍有收获但大体没有多大的变化。自从牛顿仓促地树起现代科学的旗帜，科学就勇往直前了。跟从前相比，我们正朝着更高的目标前进。而我们的探索之路将从一个简单的问题开始。

什么是空间？

## 第2章　宇宙与桶

空间究竟是人为的抽象，还是物理实体呢

一桶水能够成为一场长达300年之久的论战的核心问题的机会并不多。但是伊萨克·牛顿爵士的这桶水并不寻常。爵士在1689年记述了一个小小的实验。自那以后，这个实验对世界上最伟大的一些物理学家产生了意义深远的影响。这个实验说的是：将一只水桶装满水，然后用绳子吊起来；将绳子紧紧地拧起来，使得松手后桶能够旋转起来；拧紧后松手。起初，桶开始旋转，而其中的水处于静止状态；水的表面光洁平坦。慢慢地，随着桶的速度越来越快，其运动就会通过摩擦力缓缓地传递给桶中的水，从而使得水也开始旋转。这个时候，水的表面变成凹状，边缘高中心低，如图2.1所示。

这就是我们要讨论的实验——并不是那种使人心跳加速的问题。但是我们稍微思考一下，就会发现这桶旋转的水令人困惑不解。接下来我们就好像300年来并没有人研究过它一样，紧紧地抓住旋转的水桶这一问题，将其视为通向了解宇宙奥秘之路的关键几步。当然，理解这个问题需要一些背景知识，但是在这方面下点工夫将会是值得的。

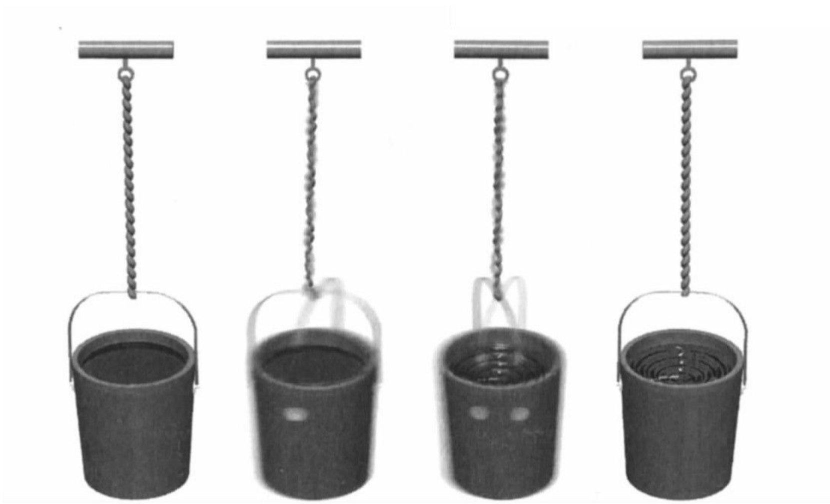


图2.1　水面开始是平的。桶刚开始旋转的时候，水面保持平稳。接下来，水也随之旋转，从而导致水面下凹。在水旋转的过程中，水面将一直下凹，即使桶减速并停止旋转的时候依然如此。

## 爱因斯坦之前的相对论

我们总是将“相对论”这个词与爱因斯坦联系起来，但事实上这一概念可以追溯到更为久远的年代。伽利略、牛顿以及其他的很多人都深知速度——物体运动的速率与方向——具有相对性。使用现代术语，我们可以这么解释：在击球手看来，投得不错的快球以差不多100千米/小时的速度飞来。但从棒球的角度看，是击球手以差不多100千米/小时的速度接近。这两种说法都是准确的，只是立场不同而

已。运动只有在相对的条件下才有意义：一个物体的速度究竟如何，只有通过与另一个物体的关系才能说明。或许你曾有过这样的经验：坐在火车中，窗外的另一辆火车与你乘坐的这辆火车相对运动。这时，你不能立刻说出究竟是哪一辆火车真的在动。伽利略在描述这一问题时使用的是他那个时代的交通工具——船。在平稳行驶的船上自然地松开一枚硬币，如同在陆地上的情况一样，硬币会落到你的脚上。从你的角度来看，你处于静止状态而水流从船边流过。这样的话，由于你并没有运动，硬币相对于你的脚的运动将与你在陆地上的情形一模一样。

当然，在某些情况下运动似乎具有某种内在性，你可以在没有外界事物作参考的情况下就可以感受并且宣称你肯定在运动。这就是加速运动，一种你的运动速度或运动方向发生改变的运动。如果你所乘坐的船突然向某一边倾斜，掉转船头，突然加速或减速，又或者是在漩涡中团团转的时候，你就会知道你正在运动。这个时候你并不需要选好某个参考点就可以知道自己在运动。即使你闭着眼睛也是如此，因为你能感觉到它。但是如果你处于速度大小和方向都不变的运动状态——匀速直线运动——你就没法知道自己是否处于运动状态。所以，你所感觉到的是运动的改变。

但是细想一下，这事有点怪。为什么运动的改变那么特别，具有某种内在的意义呢？如果运动是某种只有在比较下——通过与另一个物体对比来说明这个物体处于运动状态——才有意义的概念，那么为什么运动的改变就不是这样呢？为什么它不需要对比就有意义？事实上，它是否也是需要在对比下才有意义呢？有没有可能在我们每次提到或感受到加速运动的时候，都有某种隐含的对比在起作用呢？或许你会感到有点不可思议，但这个就是我们要追寻的核心问题，因为这个问题触及了与空间和时间的意义有关的一些深刻问题。

伽利略对运动的深刻洞察力使他确信地球本身就处于运动之中，而这却遭到宗教裁判所的憎恨。更加谨慎的笛卡儿为了避免相同的命

运，在他的《哲学原理》里采用了一种模棱两可的说法来表述他对运动的认识，不过这样显然无法躲过30年后的牛顿的审慎检查。笛卡儿认为物体对于自身运动状态的改变会自然的抗拒：静止的物体将保持静止状态直到有外力迫使其改变；以匀速沿直线运动的物体将保持匀速直线运动状态直到有外力迫使其改变。就这一说法，牛顿质疑道："静止"或者"匀速直线运动"这样的概念究竟是什么意思？静止或匀速是相对于谁来说的？静止或匀速是从谁的角度看？如果不是处于匀速运动，那么究竟是相对于谁或者说从谁的角度看不匀速？笛卡儿理顺了有关运动意义的几个方面；牛顿却发现了笛卡儿遗漏的一个关键问题。

牛顿，这个执着追求真理的人，曾经为了研究眼部解剖结构而在自己的眼眶与眼睛之间插上一根钝针；后期，在他担任造币局局长时制定了最为严酷的措施惩罚那些制造假币的人，为此超过100人被送上了绞刑架。一个像他这样的人绝不能容忍谬误或者不完备的推理。他要更进一步，于是他想到了水桶。[1]

桶

当我们放开水桶的时候，桶和其中的水会一起旋转起来，水的表面成凹状。牛顿提出的问题是：水的表面为什么会形成这样的形状？你也许会说：因为它在旋转呀，就像突然一个急转弯时，坐在汽车中的我们也会感受到汽车的压力。桶中的水也是如此，旋转的时候，水会受到来自于桶壁的压力，在这种压力下，桶壁处的水就只能向上延伸。这种解释当然说得通，但却没能抓住牛顿问题的根本意图。牛顿想要问的是水在旋转究竟是什么意思：水相对于什么旋转？牛顿那时仍在思索运动的基础，还远未来得及想明白诸如旋转这样的加速运动为何不需要与外部物体做比较这样的事情。[3]

自然而然的选择当然是将水桶当成参照物。但是经过论证，牛顿认识到这样行不通。试想一下，在桶最初开始旋转的时候，桶和水之

间一定会有相对运动，这是因为桶动起来的时候水不可能立即就动。即使水动起来了，水的表面也会保持平的状态。过了一小会后，水旋转起来了，这时水和桶之间没有相对运动了，水的表面却凹了下去。所以，如果我们将桶作为参照物的话，我们就将得到与我们所期望的完全相反的结论：有相对运动的时候，水面是平的；而没有相对运动的时候，水面却凹下去了。

其实，我们还可以将牛顿的水桶实验更进一步。桶再转一会，绳子又扭在一起（方向不同），于是桶就会慢慢减速并且最终静止下来，但是桶里的水还会继续旋转。这时，水与桶之间的相对运动与实验开始时的状况是一样的（除了顺时针旋转和逆时针旋转这一区别）。但是，水面的形状却不一样（实验开始时水面是平的，现在是凹下去的）；这正好说明了相对运动不能解释表面形状。

将桶作为水的运动的参照物这种可能性排除之后，牛顿继续大胆思考。他进一步论证道：试想一下，我们在一个深冷、完全虚无的空间——真空——中继续我们的转桶实验。这时我们不能重复完全一样的实验了，因为水的表面形状是部分依赖于地球引力的，在现在的这个实验条件下根本没有地球存在。为使我们的例子更具可操作性，我们需要一个漂浮在真空中的巨桶——就像游乐场里的摩天轮一样巨大；再想象一下，一位勇敢的宇航员，霍默，被捆在巨桶的内壁（牛顿用的并不是这个例子，他想到的是用一个绳子将两块石头绑在一起来说明问题，但要点都是一样的）。有证据表明巨桶在旋转。与桶中的水会形成凹面类似的是，霍默将会感受到被压在桶的内壁上。他面部紧张，腹部缩紧，头发向着桶壁拉伸过去。现在问题来了：在一个完全虚无的空间——没有太阳，没有地球，没有空气，没有油炸圈饼，什么都没有——究竟什么东西有可能作为巨桶旋转的参照物？因为我们想象的空间是完全虚无的，除了巨桶本身什么都没有，所以看起来可能没有任何东西可以作为巨桶旋转的参照物。但牛顿并不这么认为。

他的回答是选定终极容器作为参照物，而这个终极容器就是：空间本身。牛顿提出，我们所有人都存在于并且所有的运动都发生于一个透明的、虚无的舞台之中，而这个舞台本身就是一个真正的物理实在，他将其称之为绝对空间。[2]我们既抓不到摸不着绝对空间，也闻不到听不着绝对空间。但是牛顿宣称绝对空间的确存在。牛顿提出绝对空间是描述运动最真实的参照物。如果一个物体相对于绝对空间静止，那它就是真正的静止；一个物体相对于绝对空间运动，那它就是真正的运动。而且最重要的是，牛顿总结道，一个物体相对于绝对空间有加速度，那它就是处于真正的加速状态。

牛顿就这样以如下的方式解释了陆地上的水桶实验。实验开始的时候，桶相对于绝对空间旋转，而水相对于绝对空间静止，因此水的表面是平的。随着水的速度渐渐地接近于桶的速度，水也相对于绝对空间旋转起来了，于是水的表面就成了凹状。绳子拧紧的时候，桶开始逐渐减速，而水继续旋转——相对于绝对空间旋转——所以其表面仍然是凹的。因此，尽管水与桶之间的相对运动解释不了实验现象，但是水与绝对空间之间的相对运动就解释了这一点。空间本身就为定义运动提供了参照系。

桶只不过是个例子，论证过程本身当然是具有一般性的。根据牛顿的观点，当你系上安全带坐在车里时，你之所以能够感觉到你在运动是因为你相对于绝对空间运动；当你乘坐的飞机加速起飞时，你感到被压向座椅是因为你正在相对于绝对空间加速；当你在溜冰场中旋转起来时，感觉双臂被甩出去了是因为你相对于绝对空间加速。从另一个角度看，如果整个溜冰场旋转起来而你保持不动（假定你处于理想的无摩擦滑动状态）——那你和溜冰场之间仍然有相对运动——但是你不会感觉自己的胳膊被甩出去，因为你与绝对空间没有相对加速度。我们一直用人来打比方，为了不被一些无关紧要的细节误导，我们现在再来看看牛顿的例子。牛顿举的例子是用一根绳子拴在一起的两块石头，这两块石头组成的系统在真空中旋转。因为石头相对于绝

对空间有加速度，所以绳子会拉紧。总之，绝对空间对于运动的概念有决定性的意义。

但绝对空间究竟是什么？在这个问题上，牛顿的回应是含混不清加武断。他首先在《原理》中写道“我并不定义时间、空间、位置与运动，因为（这些）是众所周知的”，[3]从而回避了严格精确地定义这些概念。然后他又写出了那句著名的话“绝对空间，只取决于其本身性质，不需要任何外部事物为其做参考，永远保持不变并且不可移动”，也就是说，绝对空间就是永恒。这就是他的回答。但是我们也可以隐约感觉到，牛顿对简单的断言某种你不能直接看到、测量或者作用于其上的事物真的存在并肯定其重要性并不满意。他写道：

事实上，发现并且有效的区分某个特别物体的真实运动并不是一件容易的事情，因为运动发生于其中的那个不可变动的空间并不能被我们感知。[4]

这样一来，牛顿就把我们留在了一个有些尴尬的处境中。在描述物理学中最根本最重要的元素——运动时，他把绝对空间这样一个概念放在首要及核心的地位，却没有清楚地说明其定义，并且承认对于将这样重要的鸡蛋放在那么含糊的篮子里感到不快。很多其他人也都感到了这种不快。

空间困境

爱因斯坦曾经说过，要是某人用了诸如“红”，“困难”或者“失望”这样的词，我们都知道这是什么意思。但是到了“空间”这个词的时候，“它与心理体验缺乏直接的联系，在加以解释时存在着很大的不确定性”，[5]这种不确定可以追溯到久远的年代，人们为了解空间的意义而进行了不懈的努力。德谟克利特，伊壁鸠鲁，卢克莱修，毕达哥拉斯，柏拉图，亚里士多德，及其众多的追随者们多少个世纪以来在“空间”的意义上来来回回的斟酌。空间与物质之间的区别是什么？

空间是否为可以脱离物质而独立存在的客体？是否有真正虚无的空间存在呢？空间与物质彼此对立吗？空间是有限的还是无限的？

几千年来，人们对空间的哲学分析同神学的质疑不可分割。依照某些人的看法，神是无所不在的。这样的思想赋予了空间神圣的特征。这样的思想历程起源于17世纪的神学家、哲学家亨利·摩尔，有的学者认为摩尔是牛顿的导师之一。[6]摩尔相信完全虚无的空间并不存在，不过这是个完全无关紧要的观测事实。因为按照他的看法，即使空间中的物质全部被清空，空间中仍然充满着精神，因此空间永远不可能真正的空。牛顿的观点稍有不同，他认为空间中除了有实体物质外，还存在着“精神物质”。出于谨慎，牛顿认为这些精神的东西“不能阻挡实体物质的运动，就好像什么都没有一样”。[7]牛顿宣称，绝对空间，是神的感觉中枢。

对空间的哲学或宗教的思考具有一定的合理性，但正如前文爱因斯坦的审慎评价所言，这些思想在描述上都缺乏严密性。从那些思想中我们可以得到一个基本且可精确阐述的问题：我们到底是要将空间归为独立实体——就像我们对待其他事物，比如你手中的这本书那样，还是要将空间仅仅视作描述普通物质之间关系的一种语言呢？

与牛顿同时代的伟大德国哲学家戈特弗莱德·威廉·范·莱布尼茨坚信，按传统意义，空间并不存在。在谈到空间时，莱布尼茨宣称，空间只不过是用以描述事物之间相对位置的一种简单又方便的方法而已。在莱布尼茨看来，如果空间中没有物质存在，空间本身并没有任何独立意义，或者说空间并不存在。比如说英语字母表，它就是26个英文字母的顺序排列。也就是说字母表只不过是用来表明a与b相连，d后面第6个字母是j， X是u后的第3个字母，诸如此类。但如果字母不存在的话，字母表就没有任何意义——并没有独立存在的“精神字母”。字母表之所以存在是为了给字母提供字典中的顺序。莱布尼茨认为空间也是如此：空间除了在讨论物质位置关系时作为自然的语言出现外没有其他意义。在莱布尼茨看来，如果将空间中的所有物质都

拿走——空间完全是空的——空间就像没有了字母的字母表一样无意义。

莱布尼茨提出了一系列的论点支持他的所谓相对者立场。比如，莱布尼茨提出，如果空间真的是一个实体，就像某种背景物质，那么神就要在宇宙中选出一块位置来安放空间这一物质。但是，作为所有的决定都是建立在公正基础上，从不依靠随机和偶然的神，怎样才能在完全一致的虚无中为空间挑选出一块安置之地呢？毕竟，那些虚无完全一样呀。在受过科学训练出来的耳朵听来，这样的说法无异于诡辩。但是，即使去掉神学的元素，就用莱布尼茨提出的其他一些观点，我们也要面对很棘手的问题：空间在宇宙的什么位置？要是将宇宙整体——保持所有物质的相对位置不变——从左往右移动10英寸（1英寸≈2.54厘米），我们有可能知道吗？整个宇宙穿越空间这个物质时的速度是多少？如果我们根本不能探测到空间，或者对其有所作用，我们还能宣称空间真的存在吗？

就在这个时候牛顿带着他的桶插了进来，并且戏剧性地改变了论战的性质。尽管牛顿承认绝对空间的某些性质很难或者根本不可能直接探测到，但他同时也提出绝对空间的存在有可观测的效应：加速运动，就像在旋转的桶那里讨论的那样，是相对于绝对空间进行的。因此，根据牛顿的观点，凹下去的水面，就是绝对空间存在的证据。在牛顿看来，一旦有了某物存在的确定证据，不管这个证据多么的间接，论战也应该停止了。就是这样神奇的一击，牛顿将有关空间的哲学式思考转变成了科学上可验证的数据。其效应很明显。莱布尼茨被迫适时地承认“我同意一个物体绝对真实的运动同其相对于另一个物体在位置上的相对改变有区别”，[8]虽然这并不是向牛顿的绝对空间投降，但这却是对坚定的相对论者的一次重击。

接下来的200年间，莱布尼茨及其追随者反对赋予空间独立的实在性的论辩在物理学家中未能激起一丝涟漪。[9]相反，舆论明显地倒向了牛顿的空间观；建立在其空间观基础上的牛顿运动学定律占据了

舞台的中心。很明显，牛顿定律在对实验的精确描述上所取得的巨大成功使人们普遍接受了他的观点。但是，有必要提及的是，牛顿本人仅仅将自己在物理学上取得的巨大成就视为用以支持他所真正看重的伟大发现的坚固基础，这个伟大发现就是：绝对空间。对于牛顿来说，所有的一切都是为了空间。[10]

马赫以及空间的意义

我还是个孩子的时候，常常和我的父亲在曼哈顿的街上玩一种游戏。我们俩中的一人四下张望，悄悄地选定某件正在发生的事情——刚刚过去的公共汽车、落在窗台上的鸽子、某人口袋中掉落的硬币——然后换一个非常规的视角来描述整件事情，比如用公共汽车的轮子、飞翔中的鸽子，或者下落中的硬币的视角来描述这个过程。整个游戏的困难之处在于用不熟悉的描述方式，比如“我走在漆黑、圆柱形的表面上，周围满是低矮却布满花样的墙，一束粗壮的白色蔓枝从天而降”，然后另一个人需要猜出，此时是以一个热狗上的蚂蚁的视角在观察，而小摊贩正往热狗上加泡菜呢。早在我接触第一门物理课程之前的很多年，我们就不玩这个游戏了，但是这个游戏还是使我在初次遭遇牛顿定律时有些不满。

这个游戏强调的是要从不同的视角来看待这个世界，每个视角和其他的视角都是一样有效的。但是根据牛顿的说法，你当然可以自由地选择一个视角来观察这个世界，但是不同视角的地位绝不是相同的。从冰面上一只蚂蚁的角度看一位溜冰者的靴子，旋转的是溜冰场和整个体育馆；但是从看台上的观众的角度看，正在旋转的是溜冰者。两个视角看起来是等效的，好像有相同的地位，好像两者之间是彼此对称的关系。但是，根据牛顿的观点，某种视角比其他的视角更加接近于真相。因为，如果真的是溜冰者在转的话，他或她的胳膊将向外伸展；而如果真的是溜冰场在转的话，他或她的胳膊就不会向外伸展。接受牛顿的绝对空间的观念意味着必须接受绝对化的加速运动概念，特别是，要接受关于谁或什么东西真正在旋转的绝对化的答

案。我努力去想为什么会是这样。我参考的每一种资源——教科书或者老师之类的人——都同意当考虑的是匀速直线运动的时候只有相对运动才有意义；但是，我禁不住想到，为什么在这个世界中加速运动这么特别呢？为什么相对加速运动不能是考虑非匀速运动时唯一有关的内容呢？就像相对速度一样。也许在别的地方，绝对空间的存在没有什么；可是对于我，绝对空间实在太古怪了。

很久以后我才知道，在过去的几百年间，很多的物理学家和哲学家——或者大声疾呼，或者悄悄抛出——都与相同的问题斗争过。表面上看起来牛顿的桶说明了正是绝对空间使得一种视角比另一种视角更重要（相对于绝对空间旋转的人或物才是真正的旋转；其他的则不是），但这样的答案令很多仔细思考过这一问题的人很不满意。直觉上不应该有哪种视角比其他的视角“更加正确”。而且，考虑到莱布尼茨的合理意见，物体之间只有相对运动才有意义，我们不禁要问，为什么绝对空间能告诉我们哪个才是真正的加速运动，但却不能告诉我们哪个是真正的匀速直线运动？毕竟，如果绝对空间真的存在，它应当成为所有运动的基准，而不仅仅是加速运动。如果绝对空间真的存在，为什么它不能告诉我们在绝对意义上我们处于什么位置？那样我们就不需要根据另一个作为参照物的物体来确定我们的相对位置了。而且，如果绝对空间真的存在的话，为什么只能是它来影响我们（例如，我们在旋转的时候会使我们的胳膊向外伸展），而我们却无法对它产生影响？

在牛顿宣示其工作之后的上百年里，这些问题时不时成为争论的焦点。但直到19世纪中叶，奥地利物理学家哲学家欧内斯特·马赫登场之后，一个大胆而极富预见性的，并且产生了深远影响的新的关于空间的观点才出现在世人面前；而这一观点在适当的时候对爱因斯坦产生了深远的影响。

为了理解马赫的洞察力——或者更准确地说，为了用现代知识解读通常归功于马赫的原理[4]——让我们暂时回到桶的问题上。牛顿的

论证中有些古怪之处，桶的问题给我们的挑战是解释为什么有些情况下水面是平的，而另一些情况下则是凹的。为了给出解释，我们详细考虑了两种不同情形并且发现不同情形关键的区别在于水是否在旋转。于是问题来了：在引入绝对空间的概念以前，牛顿只把桶当成确定水运动与否的参照物，正如我们所见，这样的尝试带来了失败。然而，我们完全可以用其他一些物体作为参照物来确定水运动与否，比如我们就可以选择做实验的地方作为参照物——实验室的地板、天花板以及墙壁，等等。或许我们还可以在一个晴朗的天气跑到户外的开阔地做这个实验，周围的建筑、树木或者脚下的大地都可以当做是确定水旋转与否的“静态”参照物。要是我们在外太空做这个实验，我们还可以将远方的群星选为静态参照物。

于是我们就提出了下面的问题。是不是牛顿根本没理会这种情况呢？我们在生活中常常自然想到的那些相对运动，比如水与实验室、水与大地或者水与群星之间的相对运动，是不是被他漏掉了呢？这样一些相对运动能否在不引入绝对空间的情况下解释水面的形状呢？这就是马赫在19世纪70年代提出的问题。

为了更全面地体会马赫的观点，发挥一下你的想象力：你正漂浮在外太空，安静，不动，完全感受不到质量。四下一望，你可以看到远方的群星，它们也处于完美的静止状态（真正禅定的时刻）。就在这个时候，恰巧飘过来的某人抓住了你，使你团团转起来了。这时候你会发现两件事。第一件，你的手臂和腿会感受到向外的拉力，要是你不使劲的话，手脚会自然伸展开。第二件，你再盯着群星看的时候，会发现它们不再是静止的了。天际的群星组成了巨大的弧形，绕着你不停地旋转。你这一刻的体验将身体感受到的力与见证相对于群星的运动紧密联系起来。请先记住这件事，我们稍后换个场景再试一次。

再次发挥一下你的想象力：这次你浸在完全空荡荡的空间中，周围漆黑一片，没有星星，没有银河，没有地球，没有空气，总之除了

黑暗什么都没有（真正的存在主义的一刻）。这次你开始旋转时有什么感受呢？你还会感到手脚向外伸展吗？我们的日常生活经验告诉我们是这样的：每次我们从不旋转的状态（什么感觉都没有的状态）进入旋转状态的时候，都会感到肢体向外伸展。但是目前我们面临的这种状况是我们中的任何人前所未有的经历。在已知的宇宙中，总有其他物质，有的就在附近，有的远点（比如远方的群星），可以作为各种运动状态的参照物。但是，在现在的这个例子中，绝对没有任何办法可以让你通过与参照物对比来分辨什么是“旋转”什么是“不旋转”；这里没有任何其他物体。马赫想到了这一点，并且继续前进了一大步。马赫认为，在这种状况下，我们也没办法区分各种不同的旋转之间的差别。更准确地说，马赫认为，在一个完全虚无的宇宙中，旋转与不旋转没有区别——如果没有比较作基准的话根本无所谓运动或加速——所以旋转与不旋转是完全一样的。如果将牛顿的那两块用绳子拴在一起的石头放到完全空荡荡的空间中，绳子只能是松弛的。如果你在完全空荡的空间中旋转，你的手脚也不会向外伸展，你耳朵中的液体也不会受到任何影响。总之，你什么感觉也没有。

这一想法深奥而微妙。只有在心里面真正把自己想象成是在漆黑凝滞、完全虚无的空间中，你才能真正理解这一想法。这里说的完全虚无一物的空间并不像一间漆黑的房间，在房间里你的脚还能踩在地板上，慢慢适应一会之后你的眼睛还能习惯黑暗，逐渐能够感受到门缝或者窗户缝里透来的微光；我们需要想象出来的空间是真正的虚无一物，什么都没有，连地板都没有，即使慢慢适应了，你的眼睛也什么光亮都看不到。你彻底陷入宇宙的黑暗之中，没有任何物体与你为伴，没有任何物质可以参照。由于没有参照物，马赫论证道，运动以及加速这些概念都不再有任何意义，现在不再是你感受不到自身的旋转那么简单；现在的事实更加基本。在一个空无一物的宇宙中，你究竟是在安静地站着还是在匀速旋转中是完全不可区分的。[5]

当然，牛顿是不会同意这一点的。牛顿宣称即使在空无一物的空

间中仍然有空间。尽管空间无形且无法直接触摸，牛顿依然认为它可以作为衡量物体运动与否的参照物。但是别忘了牛顿是怎样得到他的结论的：他对旋转运动的思索是在一个假设下进行的，即大家所熟知的实验室中得到的结果（水的表面凹下去；霍默能感受到桶壁的压力；旋转的时候你的手臂会向外伸展；两块绑在一起飞速旋转的石头中间的绳子拉紧）在空无一物的空间中同样能够得到。这样的假设引导他去寻找在真空中可以作为运动的参照物的东西，而他所得到的结论是：那个可以用以定义运动的东西就是空间本身。马赫挑战的正是牛顿这一关键假设：马赫认为在实验室中能够得到的结果并不能在完全空无一物的空间中得到。

200年间，马赫原理第一次真正做到了挑战牛顿体系。接下来的很多年里，马赫原理持续着其对物理学界的影响力（其影响甚至超越了物理学界：1909年，弗拉基米尔·列宁在伦敦的时候著有一本哲学小册子，其中就讨论了马赫原理的有关内容[11]）。但是，即使马赫是正确的，即在完全虚无一物的空间中没有旋转的概念——这一论述排除了牛顿提出的绝对空间——地球上进行的水桶实验仍然是个未解之谜，水面的确凹了进去。不采用绝对空间的概念的话——如果绝对空间什么都不是——如何解决这一难题呢？这一问题的答案来自于对马赫推理的一个简单异议的思考。

马赫、运动及群星

我们来想象一个不同于马赫的想象——并非完全空无一物的宇宙，在这个宇宙中，有一些星星在天际闪闪发光。在这样的环境下你再来进行外太空旋转实验，这时，那些星星——即使由于距离的原因它们只有针孔般大小的微光——就可以用以标定你的运动状态。开始旋转的时候，远方的星光围绕着你转。既然远方的星光已经为你提供了一种可以看到的方式来确定旋转与否，你会觉得自身也可感受到这种区别。但是远方的星星如何才能做到这一点呢？它们的存在与否是如何开启或关闭你感知旋转的能力（或者更一般地说，感知加速与否

的能力）的呢？如果在一个只有少数距离很远的星星的宇宙中你能够感受到旋转运动，这是否意味着马赫错了呢？或者，正如牛顿所言，在一个空无一物的宇宙中你仍能感受到旋转。

对于这一异议，马赫给出了一种回答。根据马赫原理，在一个空荡荡的宇宙中，你不会感受到你是否旋转（更准确地说，根本就没有旋转或不旋转这样的概念）。另一方面，在由群星以及其他我们真实宇宙中存在的事物组成的宇宙中，旋转的时候你会感受到你的手臂及大腿受到一种向外伸展的力的作用（自己试试）。现在就是问题的关键所在了，在一个既不是完全空无一物，又没有我们的宇宙那么多的物质的宇宙中，马赫认为你在旋转时所感受到的力介于零和在我们的真实宇宙中所感受到的力之间。也就是说，你所感受到的力正比于宇宙中物质的数目。在一个只有一颗星星的宇宙中，你在旋转时所感受到的力极小。有两颗星星存在的话，感受到的力会大一点。依此类推，一直到其中的物质与我们的真实宇宙的物质一样多的时候，你在旋转时所感受到的一切会令你十分熟悉。按这种方法，你从加速运动中所感受到的力实际上是宇宙中所有物质的一种累加效果。

当然，这种观点对于所有的加速运动都成立，而不是仅仅适用于旋转。你所乘坐的飞机加速离开跑道的时候，你开的车伴着刺耳的声音刹住的时候，你所乘的电梯开始攀升的时候，马赫的想法告诉你：你所感受到的力是组成宇宙的所有物质的共同影响。如果宇宙中的物质更多，你所感受到的力也应该更大；如果宇宙中的物质更少，你所感受到的力也应该更小。而如果宇宙中没有物质存在的话，那你将什么都感受不到。所以，按照马赫的方式思考，实际上重要的只有相对运动和相对加速度。仅当你相对于宇宙中其他物质的平均分布加速的时候你才能感受到加速运动。什么其他物质都没有的话——没有任何可以参考的基准——马赫则认为没法感受到加速运动。

对于大多数物理学家而言，这是在过去150年间关于宇宙的看法中最诱人的一种。令几代物理学家深感不安的那种触不到、抓不着、

理解不了的空间，真的就是可以用来作为衡量运动的根本的、绝对的基准吗？对于很多人来说，这实在太荒谬了。至少在科学的意义上，基于某种完全感觉不到的，大大超过了我们的认知范围的，甚至可以说接近神秘的东西来理解运动，这样的做法并不可靠。但是这些物理学家们也同时受困于如何解释牛顿的水桶。马赫的见解之所以能够带来一股兴奋之情完全在于他发现了新解释的可能性，而其中空间并不是关键，马赫的回答把人们再次带回到莱布尼茨所倡导的空间的相对论者观念。在马赫看来，空间——正如莱布尼茨所想的那样——只不过是表示物体之间相对位置关系的语言。就像没有了单词的单词表什么都不是，空间也不喜欢独立存在。

马赫与牛顿

我在读大学的时候知道了马赫的想法，马赫原理真是天赐之物。终于，我们有了一个可以将所有的视角等同起来的空间和运动的理论，而所有视角之所以地位相同是因为只有相对运动和相对加速度才有意义。不同于牛顿判定运动的基准——那个不可见的绝对空间——马赫的基准是让所有人都能看见的——那就是遍布于宇宙的物质。我确信马赫原理必是正确的答案。我也发现我并不是唯一一个有这样反应的人；在我之前的很多物理学家，包括阿尔伯特·爱因斯坦，在初次了解到马赫原理的时候便对其一见钟情。

那么马赫真的是对的吗？牛顿真的陷入桶的泥潭而不可自拔以至于得到的是有关空间的孱弱结论？到底是牛顿的绝对空间真的存在，还是历史的钟摆最终摆回了相对论者的一边呢？马赫引入他的想法之后的几十年里，人们无法回答这些问题。问题的关键在于马赫的想法并不是一种完备的理论或者描述，马赫从未能指出宇宙中的物质是如何施加他所提出的那种影响的。如果马赫的观点是正确的，那么远方的星星和你家隔壁的房子是如何在你旋转的时候使你感受到旋转的呢？由于没能提出一种物理机制来实现他的想法，我们很难定量地考察马赫的观点。

以我们现代的角度看来，合理的猜测是，与马赫提出的影响力有关的可能是引力。在马赫提出他的观点之后的几十年里，这种可能性吸引了爱因斯坦的注意。在提出自己的引力理论——广义相对论时，爱因斯坦就从马赫的想法中汲取了大量的灵感。当相对论最终尘埃落定时，关于空间是否是一种物质——或者说相对论者和绝对论者的空间观点哪种正确——这一问题，以一种将以前所有看待宇宙的方式全部粉碎的形式再次出现在我们面前。

## 第3章　相对与绝对

空间究竟是爱因斯坦式的抽象，还是物理实体呢

有些发现可以回答某些问题，有些发现则更为深刻，能够以一种全新的角度提出问题，使人们发现之前的神秘之处不过是因缺乏知识而造成的误解。你可能会穷尽一生的时间——很多古人的确如此——来思考地球的边缘是什么样的，或者试图想出是谁或者何物居住在世界的尽头。但当你发现地球是圆的，你会认识到之前的神秘问题没法回答，实际上，那个问题问得并不切题。

20世纪的头20年，阿尔伯特·爱因斯坦做出了两项重大发现。每一项发现都使人类对于空间和时间的认识发生了巨大的变化。爱因斯坦拆除了牛顿建立的严格、绝对的结构，然后以一种前所未有的方式将时间和空间综合起来进而建立了自己的体系。爱因斯坦完成他的工作之后，时间和空间就成了不可分割的统一整体，空间或时间的实在性再也无法通过分别思考空间或时间来得到了。所以到了20世纪30年代末，有关空间的实体存在问题就彻底过时了；按照爱因斯坦式的重组，我们应该问的是：时空是某种事物吗？就是这一小小的修改，使得我们对于实在性的舞台的理解完全换了一种样子。

真空真的是空的吗

在爱因斯坦于20世纪的头几年编写的相对论剧本中，光才是主角。为爱因斯坦那不可思议的洞察力搭建起舞台的正是詹姆斯·克拉克·麦克斯韦。早在19世纪中叶，麦克斯韦第一次发现通过4个强大的方程，人们可以在一个严格的理论框架下很好地理解电、磁及其之间的密切联系。[1]仔细研究英国物理学家麦克尔·法拉第的工作之后，麦克斯韦写出了这套方程组。法拉第早在19世纪早期就做了成千上万次实验，研究迄今为止仍未完全搞清楚的电和磁的特点。法拉第的关键性突破在于提出“场”的概念。后来被麦克斯韦和其他科学家加以拓展延伸，“场”的概念在前两个世纪的物理学发展中产生了不可估量的影响，并且解释了我们在日常生活中所遇到的许多小秘密。通过机场安检时，你有没有注意到那台机器是怎样做到不接触你却可以探测到你是否携带有金属物品的？做过核磁共振成像（MRI）吗？一台完全在体外的机器究竟是怎样详细地绘制出体内的图像的？就算你完全不动手，指南针的针头也会自动指向北方，这是怎么回事呢？指南针这个问题与地球的磁场有关。事实上，前两个问题也可以用磁场的概念加以解释。

我见过的最好的感性认识磁场的方式就是小学课堂里的演示：铁屑在条形磁铁附近的分布。轻微的震荡后，铁屑以规则的弓形排列，起于磁铁的北极，止于磁铁的南极，就像图3.1所示。铁屑的分布就是一个直接的证据，它说明磁铁创造了一种存在于周围空间的、不可见的物质——这种物质可以对金属碎屑这样的东西有力的作用。这种不可见的物质就是磁场，根据我们的直觉，它类似于可以充满某片空间的薄雾或香气，并可以作用于磁铁物理范围之外的物体。磁场与磁铁的关系就如同战场与指挥官，或审计员与国税局：影响远在它们的物理范围之外，它允许力在场中作用于其他物体。而这也是磁场为什么被称作力场的原因。

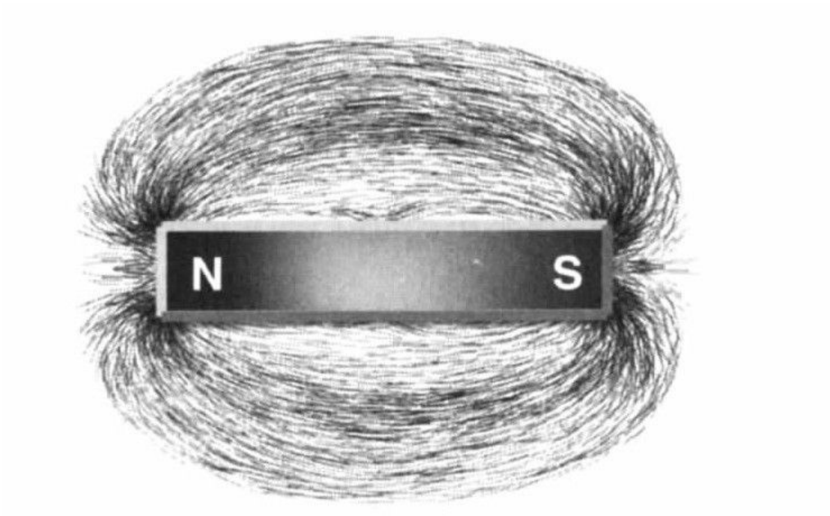



图3.1　铁屑在条形磁铁附近的分布描绘出了磁场的分布。

磁场弥漫于空间的能力使其非常有用。机场金属探测器的磁场透过你的衣服，使你带着的金属物体也发出其自己的磁场——这些磁场反作用于探测器，从而使它发出警报。MRI的磁场透过你的身体，使体内特殊的原子以适当的方式旋转并产生它们自己的磁场——然后这些磁场被探测器探测到，解码成一幅内部组织图。地球的磁场透过指南针的外壳，使指针发生偏转，指向北方，这是由于长年地球物理学过程，使地球磁场方向基本与南北极方向相符。

磁场是一种我们很熟悉的场，但法拉第还分析研究了另一种场：电场。正是由于这种场的存在，羊毛围巾发出噼里啪啦放电的声音；我们接触金属门把手后，与毛毯接触就会发出吡吡的声音；在一个电闪雷鸣下着暴雨的晚上，我们站在山顶上时会有皮肤刺痛的感觉。如果你碰巧在风雨交加的晚上看指南针，磁针偏转的方向和周围电闪雷鸣的环境就会启示我们：电场和磁场之间存在着深层次的联系——电场与磁场之间的联系是由丹麦物理学家汉斯·奥斯特首先发现的，后来法拉第又勤奋地做了很多实验对其进行进一步研究。这就像股票市场的发展会影响债券市场，反过来债券市场也会对股票市场产生影响一

样，科学家们发现，电场的变化会使附近的磁场发生变化，而磁场的变化也会造成电场的变化。麦克斯韦发现了这种联系的数学基础。因为麦克斯韦的方程表明电场和磁场之间是可以相互纠缠的，就像塔法里教的长卷发[6]互相纠结在一起那样，最终它们被命名为电磁场，电磁场可以通过电磁力作用于其他物体。

今天，我们长久地生活在电磁场的海洋中。移动电话和汽车广播在无限宽广的空间内工作着，因为电信公司和广播站的电磁场充斥着广阔的空间。无线网络连接环绕在我们身边；电脑从震荡在我们周围的电磁场——事实上，这些电磁场也穿过了我们——中采集信息形成了整个万维网。当然，在麦克斯韦时代，电磁场技术还没有充分发展起来，但是科学家们已经公认了麦克斯韦的伟大成就：麦克斯韦通过场的理论指出尽管电和磁是有区别的，但它们实际上是一种物理实体的不同方面。

后来，我们遇到了各种各样的场——引力场、核子场、希格斯场，等等——我们越来越认识到场对于现代物理学定律的形成起着十分关键的作用。但到现在，在我们讨论的领域中，关键性的下一步也归功于麦克斯韦。麦克斯韦进一步分析他的方程后发现，变化的电磁场以波的形式传播，速度为每小时670000000英里（1英里≈1.609千米）。这正是其他实验所发现的光的传播速度，麦克斯韦意识到光也属于电磁场，它可以作用于我们视网膜上的化学物质，从而使我们产生光感。麦克斯韦做出了举世瞩目的伟大发现：他将磁铁产生的力、电荷产生的力，以及在宇宙中所能看到的光联系起来——但这就提出了一个更为深刻的问题。

当我们说光速是每小时670000000英里时，经验以及前面的讨论告诉我们，如果没有参照物的话这种说法将毫无意义。有趣的是，麦克斯韦只给出了这个数值而并未提到任何参照物。这就像是某人说在北部的22英里外有个聚会，但却没有参照坐标，没有说明是哪儿的北部。包括麦克斯韦在内的很多物理学家，试图用类似于下面的方式来

解释方程中的速度：我们熟悉的波，比如海洋的波或声波，是在物质，或者说介质中传播的。海洋中的波涛是在水中传播的，声波是在空气中传播的，这些波的速度都是相对于介质而言的。当我们说声波在房间中的速度是每小时767英里时（也就是通常所说的1马赫，这里的马赫就是我们在前面提到过的欧内斯特·马赫），我们想要表明声波在空气中是以上述速度传播的。于是很自然的，那时的物理学家推测光波——电磁波——也是在某种特殊的介质中传播的，虽然这种介质从未被人探测到，但它肯定是存在的。这种看不见的传播光的物质被命名为光以太，或以太；后者是一个古老的术语，亚里士多德曾用它来描述一种可以包罗万象的神奇物质，在想象中，天国的东西就是由它做成。为了使该说法与麦克斯韦的结果一致，有人提出他的方程暗示着采用了相对以太静止的物体作为参照物。他的方程中的每小时670000000英里，就是光相对于静止以太的速度。

正如你所看到的那样，光以太和牛顿的绝对空间存在着惊人的相似性。它们都起源于提供一种参照物以定义运动的尝试；加速运动导致了绝对空间的概念，光的运动导致了光以太的概念。事实上，许多物理学家认为以太是圣灵——亨利·摩尔、牛顿和其他科学家认为充满绝对空间的圣灵——的实际替身（牛顿和他同时代的科学家曾用“以太”描述过绝对空间）。但实际上以太是什么呢？它是由什么构成的？它来自于哪里？它存在于每个地方吗？

这些关于以太的问题与几个世纪以来关于绝对空间的问题一样。但是，虽然关于绝对空间的完整的马赫式检验需要在全空的宇宙中旋转；但物理学家们却能提出可行的实验确定以太是否真的存在。比如说，当你游向迎面而来的浪花时，波浪向你移动的速度加快了；当你游向浪花的相反方向时，波浪向你移动的速度减慢了。类似的，当你穿过假设中的以太朝向或背离光波移动时，按照同样的推理，光波向你移动的速度比每小时670000000英里加快或减慢了。1887年，阿尔伯特·迈克尔逊和爱德华·莫雷测量了光速。经过一次次实验他们发

现，不管他们做什么运动，也不管光源做什么运动，光速总是每小时670000000英里。人们想出各种各样的巧妙说法以解释这个结果。有些人说，或许实验者是在不知情的情况下，在移动时拖曳以太与他们一起运动。有些人则大胆地猜测，或许实验设备穿过以太时变得不太正常，从而毁了实验。最后，直到爱因斯坦提出他革命性的理论，人们才终于弄清楚如何解释迈克尔逊莫雷实验。

相对的空间，相对的时间

1905年6月，爱因斯坦发表了一篇题为“运动物体的电动力学”的论文，彻底结束了光以太的历史。仅仅一击，它就永远改变了我们对空间和时间的理解。1905年4月和5月，经过5个星期的高强度研究工作，爱因斯坦的思想在这篇论文中最终成型，这个问题烦扰了他将近20年。还在少年时代，爱因斯坦就在考虑这样一个问题，假如你以光速奋力追赶光，那光波看起来将会是什么样子。因为你和光都以相同的速度飞快地穿过太空，你将和光保持同样的步伐。爱因斯坦的结论是，这样以你为参照物的话，光是不运动的。如果你伸出手去，就可能抓到一把不运动的光，就像你抓住从天空中飞落的雪花一样。

但问题是，麦克斯韦的方程不允许光处于静止状态——光不能看起来不动。而且很明显，没有任何可靠报告说过人能抓住一把静止的光。因此，年轻的爱因斯坦就问，我们的推理究竟是哪里发生了矛盾呢？

10年后，爱因斯坦用他的狭义相对论给了我们一个答案。虽然关于爱因斯坦的发现的理性根源有许多争论，但毫无疑问的是，他对于简易性不可动摇的信念在他发现狭义相对论的过程中起了非常重要的作用。爱因斯坦知道至少有一些实验没有探测到以太的存在。[2]既然这样，我们为什么非要试图找出实验的错误呢？相反，爱因斯坦认为，从最简单的方面思考：这些实验没有找到以太是因为以太根本就不存在。因为麦克斯韦的方程描述了光的运动——电磁波的运动——

不需要任何介质，实验和理论都得出了相同的结论：光，不像我们曾经遇到过的任何一种波，它不需要介质就可以传播。光是一位孤独的旅行者。光可以在真空中穿行。

但是我们是如何根据麦克斯韦的方程得出光速是每小时670000000英里的？如果没有以太作为基准的话，那这个速度又是从何而来的？又一次，爱因斯坦颠覆了传统，最终用简单性回答了这个问题。如果麦克斯韦的理论没有使用任何静止的参照物，那最直接的解释就是我们根本不需要参照物。爱因斯坦解释道："光速相对于任何物体而言，速度都是每小时670000000英里。"

这种说法显然相当简单，它非常符合爱因斯坦的座右铭："使一切事情尽可能简单化，除非不能更简单。"这个问题看起来有点疯狂。如果你追着一束光跑，常识告诉我们以你为参照物的话，光速比每小时670000000英里要慢。反之，如果你朝着光跑，常识告诉我们光速比每小时670000000英里要快。在其一生中，爱因斯坦总要挑战常识，这次也不例外。他有力地论辩道，不管你跑得有多快，也不管你是朝着光跑还是背离光跑，你测量到的光速将总是每小时670000000英里——不会比这多，也不会比这少。这当然就解决了困扰过少年爱因斯坦的问题：麦克斯韦的理论不允许静态的光存在，因为光永远都不会静止；不管你处于何种运动状态，你朝着光跑或是背离光跑，或者静止不动，光速都不会发生变化，它总是每小时670000000英里。但是，我们不禁要问，光为什么会有这么奇怪的现象呢？

先考虑一下速度。速度是通过某物运动的距离除以通过该距离所用的时间而得到的。它是对空间（运动的距离）的测量与对时间（该段路程的运动时间）的测量的比值。从牛顿的时代起，空间被看成是某种绝对存在，某种"与外界的任何物质无关"的存在。因此，空间的测量和空间间隔都是绝对的：不管测量空间中两个物体之间距离的是谁，只要他认真测量，所得到的答案都是一致的。虽然我们没有直接

这样说过，但牛顿宣称时间也是如此。牛顿在《自然哲学之数学原理》一书中用他以前描述空间的语言来描述时间："时间的存在和流逝与外界的任何事物无关。"换句话说，牛顿认为，存在着一个普适的、绝对的时间概念；这样的时间概念可用于任何地点，任何时刻。在牛顿式的宇宙里，不管测量某件事所用时间的人是谁，只要他认真测量，所得到的结果都应是一致的。

这种关于时间和空间的假说与我们在日常生活中的体验一致。正是以这样的常识为基础，我们才会说，如果我们追着光跑，光速看起来将会减慢。为了搞清楚这个观点，我们来想象一下巴特，他曾有一个核动力的溜冰板，他决定用来做终极挑战——追着光跑。虽然当他看到溜冰板的极限速度是每小时500000000英里时有点失望，但他决定还是做出他的惊人举动。他妹妹站在准备好的激光前，从11开始倒数（11是她的偶像叔本华最喜欢的数字）。等数到0时，巴特和激光飞奔出去。莉莎看到了什么呢？过去的每1个小时里，莉莎看到光移动了670000000英里，而巴特只走了500000000英里，这样莉莎就得出结论，光每小时都比巴特多走170000000英里。现在我们再来看一下牛顿的理论。他的观点表明莉莎关于空间和时间的观察是绝对的、普遍的，任何人做这个实验都可以得到相同的答案。对于牛顿而言，有关运动在空间和时间中的这些事实，就和2乘以2等于4一样都是客观的。按牛顿的说法，巴特会同意莉莎的看法，他会说光波每小时比他多走170000000英里。

但是回来后的巴特说，他完全不能同意这种看法了。他沮丧地说不管他做什么——不管他怎样推溜冰板——他看见的光速总是每小时670000000英里，一点也不少。[3]如果你不相信巴特，可以去看看过去100年间数以千计的设计精妙的实验。这些实验都是利用移动的光源和接收者来测量光速，而所有的结果在很高的精度上支持巴特的观测事实。

为什么会这样呢？

爱因斯坦指出，这个答案符合逻辑，而且是对我们到目前为止的讨论内容的深度拓展。巴特对距离和时间间隔的测量——巴特用来搞清楚光比他快多少所需要的信息——一定不同于莉莎对距离和时间间隔的测量。想一下，因为速度无非是距离除以时间，对于巴特而言，在光比他快多少这个问题上，没有理由得出一个不同于莉莎的结果。因此，爱因斯坦得出结论：牛顿关于绝对空间和时间的观念是错误的。爱因斯坦意识到相对于彼此运动的实验家们，就像巴特和莉莎，在空间和时间的测量上，是不会得出相同的结果的。关于光速的这种令人费解的实验数据只能通过空间感和时间感上的不一样来解释。

狡猾但不恶毒

空间和时间的相对性是一种令人惊奇的结论。我已经了解它25年了，但即使是现在，每当我坐下来静静地思考它时仍会感到迷惑。从光速恒定这一乏味的说法中，我们可以得出这样的结论：空间和时间是针对旁观者的角度而言的。我们每个人都有一个自己的时钟；对于时间的流逝，我们每个人有自己的认识。每个人的时钟都一样精确，但若我们相对于其他人运动的话，这些时钟就会不一致。它们是不同步的；用它们测量两个给定事件之间流逝的时间，不同的时钟测得的量是不一样的。对于距离也是一样的。我们每个人都有自己的准绳，对于空间中的距离，我们每个人有自己的认识。每个人的准绳都同样精确，但当我们相对于其他人运动的时候，这些准绳就会不一致；用它们测量两个定点之间的距离，不同的准绳测得的量是不一样的。如果空间和时间不是这样的话，光速就不恒定了，它将取决于观测者的运动状态。但光速是恒定的；空间和时间确实是这样的。空间和时间以精确的方式互相补偿，从而使得人们测量光速时总是得到同样的结果，无论观测者的速度怎样都是如此。

定量上精确地找出空间和时间的测量结果究竟有何不同是非常棘手的，但所需要的却只是高中水平的代数而已。使爱因斯坦的狭义相对论富于挑战性的并不是数学的深度。而是由于他观点上的与众不

同，且不符合我们的日常生活经验。但只要爱因斯坦参透了关键的一点——需要打破200多年来牛顿关于空间和时间的观点——完善整个理论的细节之处就将没有任何难度。在每次测量光速都能得到同样结果的前提下，爱因斯坦能够精确地算出两个不同观测者在空间和时间的测量上的差别究竟有多大。[4]

为了更深刻地理解爱因斯坦的发现，让我们再来想想巴特，他曾经激情满怀地拿出了他那最高时速可达65千米的滑板。如果他向着北方高速运动——朗读、吹口哨、打哈欠，或者偶尔在马路旁张望，然后消失在往东北方向去的高速路，那么他朝北运动的时速将小于65千米。原因很明了。刚开始，他所有的速度都是贡献于向北的运动，但是当他转向时，一部分速度贡献给了向东的运动，只留下一部分贡献于向北的运动。这个相当简单的例子实际上帮助我们抓住了狭义相对论的核心内容。以下是解释：

我们习惯于认为物体可以穿越空间，事实上另一种运动也非常重要：物体也可以穿越时间。举个例子来说，腕上的手表、墙上的时钟都显示着时间在滴滴答答地溜走，这就意味着你和你周围的一切事物都在不断地穿越时间，不停地从一秒到下一秒。牛顿认为穿越时间的运动完全不同于穿越空间的运动——他认为这两种运动之间不存在什么联系。但爱因斯坦却发现它们紧密相连。事实上，狭义相对论的革命性发现在于：当你注视某物，比如一辆静止的车时，以你作为参照物的话它是静止的——没有穿越空间，也就是说——这辆车的所有运动仅是穿越时间。车、司机、马路、你以及你的衣服都在同时穿越时间：一秒接着一秒，在滴答声中均匀地溜走。但是如果车开走了，它的一部分穿越时间的运动将转换成穿越空间的运动。这正如巴特将一部分向北的运动转换成向东的运动，从而使得向北运动的速度减慢了一样；车子的一部分穿越时间的运动转换成了穿越空间的运动，从而使车穿越时间的运动的速度减慢了。也就是说，汽车穿越时间的运动减慢。因此，相比于静止的你我而言，运动中的汽车和司机所感受的

时间流逝要慢一些。

简而言之，这就是狭义相对论。事实上，我们可以更加精确的一步步地描述这个过程。由于设备问题，巴特只好把时速限制在65千米。这在整个过程中是非常重要的，因为当他转向东北方向时，如果可以加速，它就可以弥补在转速中损失的速度，从而维持朝向北的速度。但是由于这个限制，不管他如何努力加速，他的整个速度——北部和东部两个方向的合速度——仍然保持在最多每小时65千米。这样一来，当他把方向向东转一点，当然会减慢向北的速度。

狭义相对论提出了一个适用于所有运动的简单原则：任何物体穿越空间和穿越时间的合速度总是精确的等于光速。仅凭直觉你可能不会接受这种观点，因为我们都已经习惯了只有光才能以光速运行。但是这个众所周知的说法指的只是穿越空间的运动。而我们现在的讨论与此有关，但更加复杂：一个物体穿越空间和时间的合速度。爱因斯坦发现一个关键的事实：这两种运动总是互补的。当你刚才注视的那辆静止的车开走时，真正发生的情况是穿越时间的运动的一部分速度转换成了穿越空间的运动速度，但保持合速度不变。这样的转换无疑意味着汽车穿越时间的运动将会减慢。

我们再来看看刚才那个例子，当巴特以每小时500000000千米的速度运动时，如果莉莎能看到巴特戴的手表时，她将会发现巴特手表的运转速度是她自己戴的手表的2/3。换句话说，莉莎的手表每过3小时，巴特的手表将只过2小时。他在空间的快速运动大大减慢了他穿越时间的速度。

而且，当穿越时间运动的所有速度都转化为穿越空间运动的速度时，将达到穿越空间的最大速度，即光速——这样我们就能理解为什么不可能以大于光速的速度在空间运动。光在空间总是以光速运动，光的特别之处就在于它总能完成这样的转换。就像只向东行驶中的汽车不需要有向北的速度，光所有速度都贡献于空间运动，而在时间上

无运动！当物体在空间中以光速运动时，时间就停止了。如果光粒子戴有表的话，那这个表将完全不动。光实现了庞塞·德·莱昂和宇宙工业的梦想：它没有年龄。[5]

说得更清楚一点就是，当速度（穿越空间）只比光速小一点时，狭义相对论的效应最明显。但我们所不熟悉的，穿越空间和穿越时间的运动之间的互补性，总是适用的。速度越小，与相对论之前时代的物理——也就是说，物理常识——之间的偏差就越小；但偏差虽然小却必然是存在的。

真的就是这样。这并不是巧妙的文字游戏、诡辩或者心理上的幻象。宇宙实际上就是这样运行的。

1971年，约瑟芬·海福乐和理查德·基廷乘坐一架商业喷气机飞越了整个世界，飞机上带有当时技术水平所能达到的最精确的铯束原子钟。当他们把飞机上的钟和静止在地面上的钟相比较时，发现飞机上的钟比地上的钟走得慢。尽管差别很微小——1秒钟的几百亿分之一——但却与爱因斯坦的发现一致。我们不可能找到比这更能说明问题的证据了。

1908年，传言有更新更精确的实验发现了以太存在的证据。[6]如果这些传言是真的话，就将意味着存在着绝对的静止标准，爱因斯坦的相对论是错误的。听到这个传闻，爱因斯坦回答道："上帝是狡猾的，但他并不恶毒。"窥探大自然的奥秘，弄清空间和时间的本质是一项非常具有挑战性的工作，它几乎打败了除爱因斯坦外的每个人。但是，允许这样一个令人惊奇而优美的理论存在，但又不让它与宇宙有任何联系，无疑是很恶毒的。相信这些实验的话，爱因斯坦的理论就不复存在了；但他根本不相信这些实验。爱因斯坦的信心决非空中楼阁。这些实验最终被认定是错误的，光以太从此便从科学发现中销声匿迹了。

## 但是桶呢

对光而言，这当然是一个很利落的理论。理论和实验都认为光的传播不需要介质，无论光在什么样的介质中传播，也不论人们如何观测，光的速度总是恒定不变的。所有的观测点的地位都是一样的，没有绝对的或首选的静止标准。很好。但是我们前面提到的水桶实验中的桶呢？

记住，虽然许多人都由于信任了牛顿的绝对空间，而把光以太看做物理实体，但它却与牛顿为什么引进绝对空间没有关系。相反，在与一些诸如旋转的桶这样的加速运动奋战之后，牛顿没有选择，只能引进一些无形的背景以使运动可以被明确的定义。对付了以太，却没能对付得了水桶，那么爱因斯坦和他的相对论是如何处理这个问题的呢？

好吧，说真的，在狭义相对论中，爱因斯坦的关注点集中于一种特殊的运动：匀速运动。直到10年之后的1915年，爱因斯坦才通过他的广义相对论，得以全面的把握更为普通的加速运动。虽然如此，爱因斯坦和其他人还是一而再地用狭义相对论来思考旋转运动的问题；他们认为，正如牛顿而不是马赫认为的那样，即使在一个全空的宇宙里，你也可以感受到来自旋转的外推力——霍默会感受到旋转的桶的内壁的压力；旋转的石头之间的绳将由于拉直而富有张力。[7]没有了牛顿的绝对空间和时间，爱因斯坦应该如何解释这一切呢？

答案令人惊奇。尽管名字是相对，但爱因斯坦的相对论并没有预先声明一切事物都是相对的。狭义相对论确实声称某些事情是相对的：速度是相对的；空间之间的距离是相对的；持续时间是相对的。但狭义相对论实际上引进了一种全新的、颠覆性的绝对概念：绝对时空。对于相对论而言，绝对时空是绝对的，正如对于牛顿而言，绝对的空间和绝对的时间是绝对的；部分由于这个原因，爱因斯坦并不建议使用或者特别喜欢“相对论”这个名字。相反，他和其他物理学家建

议用不变性理论这个名字，以便强调这样的理论，究其本质，乃是与那些对于每个人都一样的事物，而不是相对的事物有关的理论。[8]

绝对时空是水桶故事非常重要的下一章，这是因为，即使在定义运动时放弃所有的物质基准，狭义相对论的绝对时空还是能提供某些东西，使得物体可以相对于它们加速运动。

雕刻空间与时间

我们来看一个例子，想象一下玛吉和莉莎，为了追求生活质量，一起注册了伯恩斯学院开设的有关城市重建的拓展课程。她们首次的作业是，重新设计斯普林菲尔德的大街小巷，而且要服从两个要求：第一，街道的网格构成必须使翱翔核纪念碑恰巧位于网格中心，即在第五大街和第五大道交界处；第二，设计必须用100米长的大街，100米长的大道要垂直于大街。就在上课之前，玛吉和莉莎对比她们的设计，意识到一些事情完全搞错了。在合理地设计坐标图以使纪念碑位于中心后，玛吉发现Kwik-E-Mart位于第八大街和第五大道，核电厂位于第三大街和第五大道，如图3.2（a）。但是在莉莎的设计中，位置完全不同：Kwik-E-Mart位于第七大街和第三大道的拐角处，核电站位于第四大街和第七大道，如图3.2（b）。很显然，有一个人犯了个错误。

经过一番思考之后，莉莎意识到是怎么回事了。她和玛吉都是对的。她们只是为她们的大街和大道的坐标图选择了不同的方位。玛吉的大街和大道垂直于莉莎的；她们的坐标图相对于彼此旋转了；他们把斯普林菲尔德切割成两种不同形式的大街和大道[如图3.2（c）]。这个课程很简单，但是很重要。关于如何把大街和大道组成斯普林菲尔德存在着一定的自由性。没有“绝对的”大街和“绝对的”大道。玛吉的选择和莉莎的一样有效——或者说其他可能的方向都是有效的。

当我们把时间画进图片里时也请记着这个观点。我们思考时习惯

于把空间作为宇宙的舞台，但在某段时间内物理过程发生在空间的某些区域。比如说，想象一下傻猫和坏鼠[7]正在进行一场决斗，如图3.3（a），发生的事件正按时间顺序以旧时代翻页相册的形式记录下来。每一页都是一个“时间片”——就像电影胶片中静止的每一帧——它显示了在某一时刻的某一区域发生了某件事情。我们翻到不同的页数就可以看到在不同的时刻发生了什么。[8]（当然，空间是三维的，相册是二维的，但我们可以把思维和画册简化一下。）理清一下术语，在一段时间内的空间区域被称为时空区域；你可以认为时空区域是某一段时间内某个空间区域内发生的所有事情的记录。

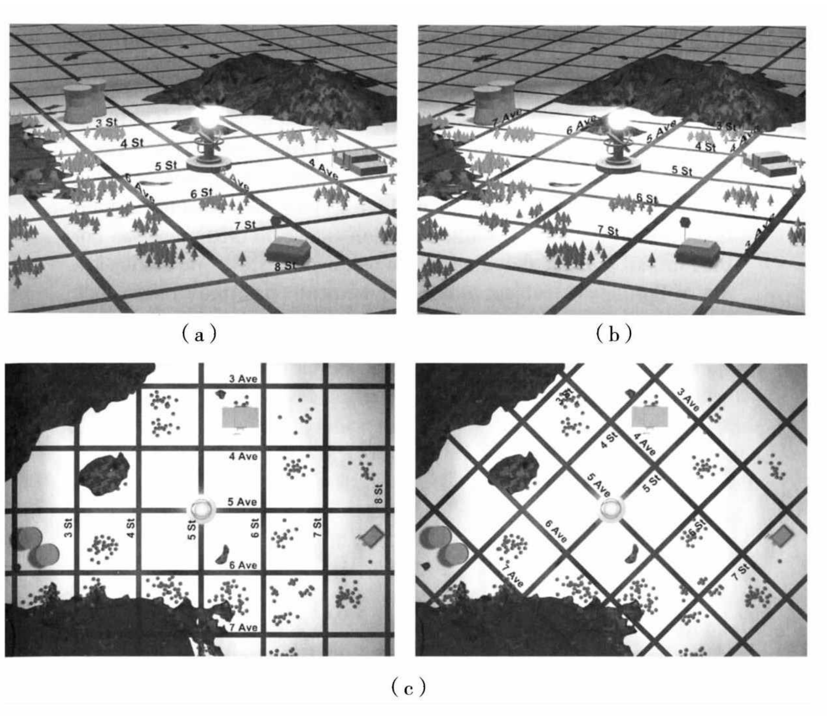



图3.2　（a）：玛吉的街道设计。（b）：莉莎的街道设计。（C）：玛吉和莉莎的街道林荫大道设计全景。她们的坐标图方向旋转了。

现在，我们来看看爱因斯坦的数学教授赫曼·闵科夫斯基的洞察力（他曾把他的年轻学生叫做一只懒狗），他把时空区域看做实体：把

完全的翻动画册看做拥有自己版权的物体。如图3.3（b）想象一下我们扩展了翻动画册的装订，就像图3.3（c），所有的页数都是完全透明的，这样你就会发现一本包含了在某个给定时刻发生的所有事情的书。从这个角度来看，这些页可以被看做是提供了一种组织模块内容的便利方式——即组织时空事件。就像大街大道坐标图通过标出大街和大道地址，可以很轻易地帮助我们使定位具体化一样，把时空板块分割成一页页可以使我们很轻易地具体化事件（坏鼠射击，傻猫被打，等等），通过给出事件发生的时间——事件发生的那一页——事件发生的具体地点在那一页有具体的描述。

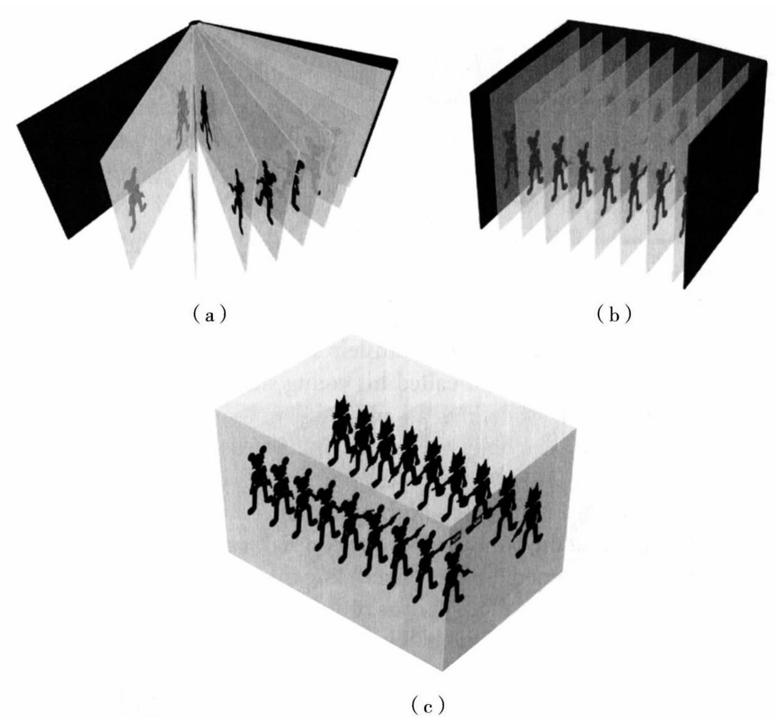


图3.3　（a）：决斗的翻页相册。（b）：扩展装订的翻页相册。（C）：包含决斗的时空模块。页码或“时间片”在模块中组织了事件。时间片之间的空间清晰可见；它们并不意味着时间是分离的，这个问题我们稍后再述。

关键在于：就像莉莎意识到把空间区域分割成街道的等效方式不

止一种，爱因斯坦意识到把时空——图3.3（c）中那样的时空条——分割成不同时刻不同区域的等效方式也不止一种。图3.3（a）、（b）、（c）中的页——再说一遍，每一页代表一个时刻——所画出的只是许多种可能的分割方式中的一种。这听起来只比我们对空间的直观感受拓展了一点点，但这一点点却是扭转我们几千年来固有的最基本直觉的基础。1905年以前，人们都认为时间的流逝对每个人来说都是一样的；大家发生在哪一时刻都会有相同的看法；因此，对于时空画册的某一页上发生了什么，大家都会有相同的看法。但是当爱因斯坦意识到相对运动的两个观测者的时钟不同时，所有的一切都变了。相对于彼此运动的时钟不再同步，因此有不同的同时性概念。图3.3（b）中的每一页，只是某一个观测者按他或她自己的时钟上的某个时刻记录下来的发生在空间中的所有事件。而相对于第一个观测者运动的另一个观测者将会发现，某一页上的所有事件并非同时发生。

这就是同时性的相对论，我们可以直观地感受到它。想象一下傻猫和坏鼠，手中都拿着手枪，对峙在长长的正在移动的火车两端，一个裁判在车上而另一个在月台上。为了使决斗尽可能的公平，所有人都同意放弃三步规则，取而代之的是当一小排火药在他们中间爆炸时，决斗者将开始动手。第一个裁判，阿布，点燃了火药抛向空中，然后返回来。当火药发亮爆炸时，傻猫和坏鼠开始开火决斗了。由于傻猫和坏鼠离火药相同的距离，阿布认为闪亮的一刹那发出的光到达他们的时间是相同的，所以他就举起了绿旗声明这个决斗是公平的。但是另一个站在月台上的裁判马丁，抱怨这是不公平的决斗，他认为傻猫比坏鼠先看到爆炸发出的光信号。他解释说因为火车向前开，坏鼠是朝向光前进的，而傻猫是远离光而去的。这就意味着光到达坏鼠不用走那么远，因为坏鼠自己就会向光靠近；而光到达傻猫需要走得更远一些，因为傻猫会远离光运动。但是因为从任何一个人的角度看光速都是不变的，所以马丁认为光需要更长的时间才能到达傻猫，因此，这样就使决斗不公平了。

谁是正确的，阿布还是马丁？爱因斯坦给出的答案出人意料：他们都对。虽然两个裁判的结论不同，但每个人的观测和推理都没有错误。就像球棒和球，它们对于事件顺序有各自不同的视角。爱因斯坦令人震惊的发现在于，由于各方视角不同，因而导致各方对同时发生的事件会做出不同但同等有效的解释。当然，就日常的速度如火车的速度而言，这个差别是非常小的——马丁认为傻猫看到光的时间比坏鼠要慢万亿分之一秒——但要是火车开得更快，接近光速，那时间上的差异就会变得重要起来。

想想这对于时空画册意味着什么。由于相对于彼此运动的观测者对同一时间发生的事情达不成一致，所以每个人把时空条切成片的方式就会不同——每一片包含的是对某个观测者而言，在某一时刻发生的所有事情。相反，相对于彼此运动的观测者以不同却同样有效的方式把时空条分割成页，分割成时间片。莉莎和玛吉从空间中发现的道理，正是爱因斯坦从时空中发现的。

调调角度

街道坐标图和时间片之间的类比可以进一步深究。就像莉莎和玛吉的设计由于坐标的旋转而不同，阿布和马丁的时间片，他们的翻页相册页——包含了时间和空间——也由于旋转而不同。这在图3.4（a）和3.4（b）中阐释得很清楚，马丁的时间片相对于阿布的发生了旋转，这使得他认为决斗是不公平的。细节的不同在于玛吉和莉莎的设计中的旋转角度只是一个设计上的选择，阿布和马丁的切片之间的旋转角度是由他们的相对速度决定的。不用花多大的力气，我们就能弄清楚是什么原因。

想象一下，坏鼠和傻猫和解了。他们不再射击对方，只是想确保火车前面和后面的钟完全同步。由于他们距离火药的距离是相同的，他们就进行了下列的计划。他们同意把他们的钟都调到中午，就像他们都看到火药爆炸发出的光一样。从他们的角度来看，光运行相同的

距离到达他们，由于光速是恒定的，因而光同时到达他们。但是就像之前的推理，马丁和其他在月台上的人都看到坏鼠是朝着光走去而傻猫是远离光而去的，因此坏鼠看到光的信号要比傻猫早一点。月台上的观测者们也因此得出结论：坏鼠把钟调到12：00的时间要比傻猫早，所以坏鼠的钟比傻猫的要快一点儿。举个例子，对于一名马丁这样的月台上的观测者来说，当坏鼠的钟是12：06时，傻猫的钟可能只有12：04（相差的数值取决于火车的速度和长度；火车越长速度越快，差异就越大）。但是，从在火车上的阿布和其他人的角度来看，坏鼠和傻猫根本就是同时进行这一动作的。虽然很难接受，但这并不矛盾：处于相对运动中的观测者并未在同时性上达成一致——即他们对于同一时间发生的事情并没有达成一致。

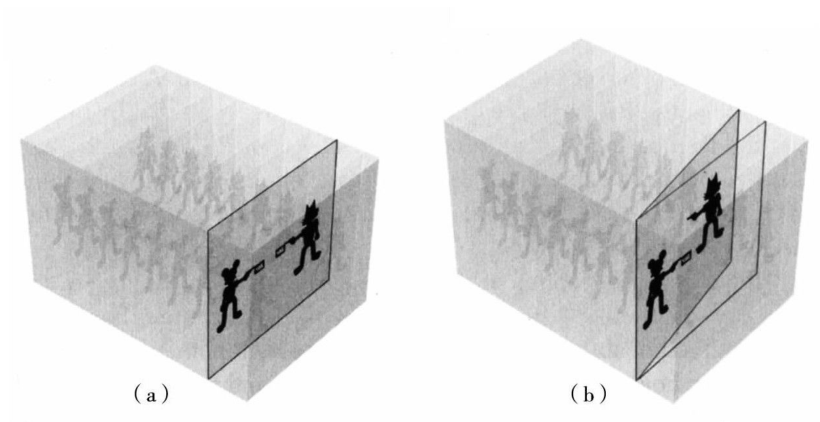



图3.4　（a）：阿布的时间片；（b）：马丁的时间片；他们两人处于相对运动中。他们穿越时间和空间的切片由于旋转了一个角度而不同。根据火车上的阿布所言，决斗是公平的；根据站台上的马丁而言，决斗是不公平的。两种观点都同样有效。在（b）中，着重强调了他们穿越时空的切片的不同角度。

这就意味着从火车上人的角度看，画册中的一页，里面包含了他们认为是同时发生的所有事情，比如坏鼠和傻猫同时调整钟；但从月台上的观测者看来，那一页中的事件却应当属于不同的页（在月台上的观测者看来，坏鼠要比傻猫早调钟，因此从月台上的观测者的角度看这些事件应该在不同的页上）。从火车上的观测者的角度来看发生

在单独一页上的事件，对于月台上的观测者而言却是发生在不同的页上。这就是为什么马丁和阿布在图3.4中的时间片相对于彼此发生了转动：从一个观测者的角度看属于同一时间片上的事情，从另一个观测者的角度看就可能属于不同的时间片。

如果牛顿关于绝对空间和绝对时间的观点是正确的，那么大家将认同单独的一张时空片。每一片将代表绝对时间中某一特定时刻的绝对空间。但世界就是这样运转的，从牛顿式的僵化到爱因斯坦新发现的弹性，这样的转变使我们的看法发生了变化。我们不再把时空看做一本不可改变的翻页册，有时有必要把它看做一块巨大的新鲜面包。这就代替了构成一本书的固定页数——牛顿时代固定的时间片——见图3.5（a），你可以从不同的角度把面包切成平行的切片。从观测者的角度来看，每一片面包代表着某一时刻的空间。但是如图3.5（b）所描述的那样，另一个相对于该观测者运动的观测者，将会从不同的角度来切时空面包。这两名观测者的相对速度越大，他们各自切片的角度差就越大（就像在注释中解释的那样，光速设定的速度极限，使这些切片的最大旋转角度为45度[9]），观测者在同一时间报告的事件的差异就越大。

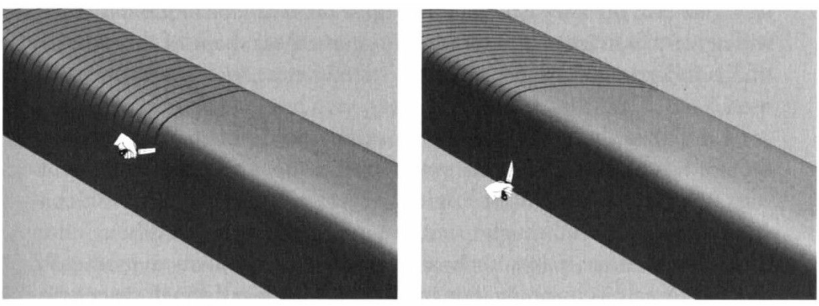



图3.5　就像一块面包可以从不同的角度切成片一样，时空模块可以被处于相对运动中的观测者从不同的角度切成“时间片”。相对速度越大，角度就越大（由于任何速度都不可能超越光速，最大角度只能是45度）。

桶，狭义相对论的观点

理解时间和空间的相对性需要我们思想上的剧变。但有重要的一点，前面提到过但现在用面包来讲，我们不应该忘记：并不是相对论中的每一件事情都是相对的。即使你我想以不同的方式切割一块面包，我们仍有可以达成一致的东西，那就是面包的整体性。虽然我们的切片有可能不同，但如果我们同时把切片组合起来，我们将得到相同的面包。为什么会是这样呢？因为我们要切的是同一块面包。

同理，在连续的时间内所有空间切片的完整性，从任意观测者的角度看（见图3.4），都会保证得到同样的时空区域。不同的观测者可以用不同的方式来切割时空区域，但时空区域本身，就像面包一样，有独立的存在性，因此，虽然牛顿肯定是错误的，但他的直觉——总有一些东西是绝对的，总有一些每个人都会认同的东西——不会完全被狭义相对论否定。绝对的空间是不存在的，绝对的时间也是不存在的。但根据狭义相对论，绝对的时空是存在的。了解了这一点，让我们再来看看桶吧。

在一个空的宇宙中，桶是相对于什么旋转呢？根据牛顿的观点，答案是绝对空间；根据马赫的观点，讨论桶的旋转是没有意义的；根据爱因斯坦的狭义相对论，答案是绝对时空。

为了理解这一点，我们再次来看看前面提到的斯普林菲尔德的街道设计图。玛吉和莉莎对于Kwik-E-Mart和核电站在街道中的地址没有达成一致，因为她们的坐标图相对于彼此旋转了。即便这样，先不考虑每个人如何设置坐标图，有一些东西她们是一致同意的。打个比方来说，为了提高午饭时间工人的效率，从核电厂到Kwik-E-Mart的地上画一条直线，玛吉和莉莎就这条线穿过几条大街和几条大道不会达成一致意见，如图3.6所示。但他们都会在这条线的形状上达成一致：必须是一条直线。这条线的几何形状是独立于一个人所使用的街道等特殊坐标的。

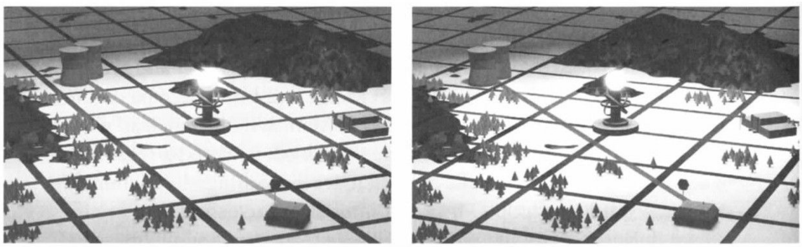


图3.6　不论采用哪一种设计图，在该例子中大家都认为轨迹的形状为直线。

爱因斯坦意识到对于时空也存在类似的问题。即使两个相对运动的观测者以不同的方式切割时空，他们仍然有达成一致的地方。拿最初的例子来看，想象一条不只穿过空间，实际穿过时空的直线。虽然时间的轨线我们不熟悉，但片刻的思考却能解释其意义。由于一个物体穿过时空的轨线是直的，则该物体不仅穿过空间的线是直的，它穿过时间的轨迹也应该是直的；这样的话，它的速度和方向都不变，因此它以恒定的速度运动。虽然不同的观测者以不同的角度切割时空面包，他们就时间流逝了多少以及一条轨线上的两点之间的距离是多长无法达成共识，但像玛吉和莉莎这样的观测者总会一致同意穿过时空的轨线是一条直线。就像到Kwik-E-Mart的轨线的几何形状是独立于街道的坐标图一样，时空中的轨线的几何形状也独立于时间片的选择。[10]

这一认识简单却关键，因为狭义相对论正是利用它，提出了一个关于判断某物是否加速的绝对标准，而这一标准是所有的观测者，不论他们的相对的固定速度是多少，都会同意的标准。如果物体穿越时空的轨迹是一条直线，就像图3.7所示的那个静静坐在那里的宇航员所留下的轨迹（a），那它就没在加速。如果一个物体穿越时空的轨线是直线以外的其他形状，则它是加速的。举个例子，要是宇航员开启引擎，像图3.7中的宇航员（b）那样螺旋运动或是像宇航员（c）那样以越来越快的速度朝着外太空飞行，他穿过时空的轨迹线就会弯曲——加速的证据。因而，有了这些新的认识，我们就可以明白：时

空轨迹线的几何形状为判断物体是否加速提供了绝对的标准。时空，而不是空间，提供了基准。

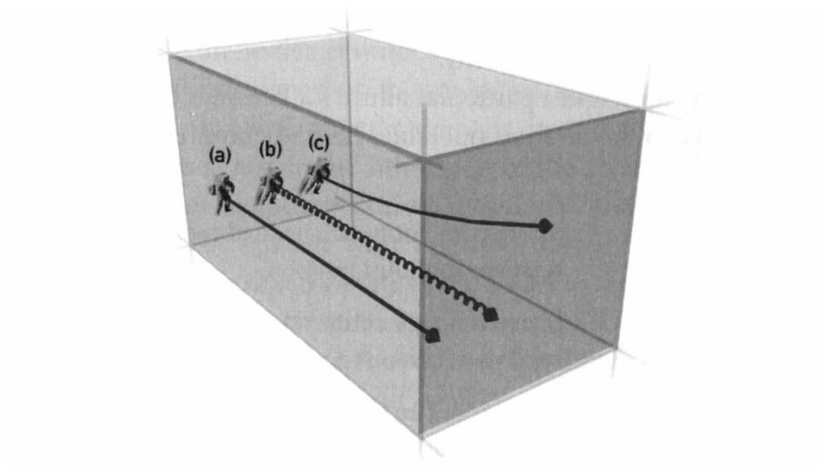



图3.7　3个宇航员通过时空模块的路径。宇航员[图（a）]没有加速因而穿越时空的轨迹是一条直线。宇航员[图（b）]在上空反复盘旋因而穿越时空的轨迹呈螺旋形。宇航员[图（C）]加速飞向深空，因而穿越时空的轨迹线发生了弯曲。

从这个意义上来讲，狭义相对论告诉我们时空本身就是加速运动的最终仲裁者。时空为像旋转的桶一样的物体提供了背景，有了这样的背景，即使在空的宇宙中我们也可以说旋转的桶在加速。从这个角度来看，钟摆又回来了：从相对主义者莱布尼茨到绝对主义者牛顿再到相对主义者马赫，现在又到爱因斯坦，他的狭义相对论又一次展示了实在性的舞台——应该是时空，而不单是空间——足以为运动提供最终的判断标准。[11]

### 引力和古老的问题

到了这里你可能会以为，我们已经讲完了桶的故事，马赫的观点不再被信奉，而爱因斯坦激进地升级了牛顿关于空间和时间的绝对概念并已经得胜。虽然事实更加奥妙并且更有趣。但如果你对迄今为止我们发现的新观点感兴趣，在进行到这一章的最后一部分之前你可能

需要休息一会儿。表3-1的总结将帮助你更新自己的记忆。

表 3－1　　关于空间和时空性质的不同立场的总结

| 代表人 | 观　　点 |
|---|---|
| 牛顿 | 空间是一个实体；加速运动并非是相对的；绝对论主义者的立场 |
| 莱布尼茨 | 空间并非一个实体；运动的所有方位都是相对的；相对论主义者的立场 |
| 马赫 | 空间并非一个实体；宇宙中加速运动是相对于平均质量分布而言的；相对主义者的立场 |
| 爱因斯坦（狭义相对论） | 空间和时间都是相对而言的；时空是一个绝对的实体 |

好。如果你读了这些话，我想你已经为了解时空观念的关键性的下一步打好了基础，这一步很大程度上正是被欧内斯特·马赫促进的。尽管狭义相对论认为——不像马赫的理论——即便在一个空的宇宙中你仍可感受到旋转的桶的内壁的压力，以及连接两个旋转的石头的绳子被拉紧的张力，爱因斯坦仍然深深地着迷马赫的观点。他意识到，对这些观点的严肃思考要求将这些思想进一步拓展。马赫并没有给出一个框架，以说明遥远宇宙中的星球和其他物质，是怎样在你旋转时使你的胳膊受到向外张开的力的，以及你在旋转的桶内感受到的内壁压力究竟有多大。爱因斯坦开始怀疑是否存在某种与引力有关的机制。

这种想法对爱因斯坦而言有一种特殊的吸引力，因为在狭义相对论中，为了使分析易于理解，他完全忽略了引力。或许，他思考着，存在一种更加健全的理论，既包含狭义相对论又包含引力，并将带来一种完全不同于马赫的观点。或许，他猜测道，对狭义相对论做包括引力的推广可能会告诉我们，物质——不管是远方的还是近处的——将决定我们加速时所感受到的力。

爱因斯坦还有另一个需要将注意力转移到引力上的更加充分的理由。他意识到狭义相对论其核心论断是光速是最快的，与牛顿关于引力的普适定律直接矛盾，而牛顿的定律在过去的200年中曾做出过里程碑式的贡献，比如精确地预测了月球、行星、彗星以及在空中运行的其他天体运动。牛顿定律尽管从实验上来说是成功的，但爱因斯坦意识到根据牛顿的定律，引力的影响无处不在——从某地到另一地，从太阳到地球，从地球到月球，从任何一个地方到另一个地方——且是瞬时产生，这意味着它比光速还快。因而与狭义相对论直接相矛盾。

为了形象地说明这个矛盾，想象一下你拥有一个令人真正失望的夜晚（家乡的球类俱乐部解散了，没有人记得你的生日，有人正在吃着最后一块芝士），你需要一点时间单独待一会儿，于是你驾着家里的轻舟午夜出航。当月亮悄悄爬到头顶时，水涨潮了（月球的引力作用造成的），皎洁的月光反射在起伏的波浪上似在浪尖舞蹈。此时此刻，你的夜晚似乎还不是最令人气馁的，意想不到的其他星系的敌人破坏了这美好的一切，他们轰击月亮，一下子就把它弄没了。月球的突然消失肯定是令人奇怪的，但如果牛顿的引力定律是正确的，刚才的那一幕将展示一些更加奇怪的东西。牛顿的定律预测水将从高潮开始退下去，由于月球引力的消失，一秒半以前还看得见的月亮突然从天空消失了。就像一个抢跑的赛跑选手，海水却是在一秒半前就开始退潮。

根据牛顿定律，原因是这样的，在月球消失的一刹那它的引力效应也随之立即消失了，没有了月球的引力，潮将很快退去。由于光花一秒半的时间才能够从月球传到地球，因而你不会立即看到月球消失。在这一秒半中，虽然皓月当空，但潮水正在退去。因此，根据牛顿的思路，引力可以比光先影响我们——引力比光快——关于这一点，爱因斯坦认为它肯定是错误的。[12]

因此，在1907年左右，爱因斯坦开始沉迷于建立一个新的引力理

论，它至少要像牛顿的理论一样精确，而且又不与他本人的狭义相对论相矛盾。这无疑是一个挑战。爱因斯坦那惊人的智力终于遇上了对手。从那时起，爱因斯坦的笔记本上写满了各种半成形的想法，几乎迷失在小小错误诱导的长长的迷途中，每次惊喜的以为就要解决该问题时又遇到了另一个错误。终于，到了1915年的时候，爱因斯坦看到了胜利的曙光。虽然爱因斯坦在一些关键的地方得到了数学家马塞尔·格罗斯曼的巨大帮助，但广义相对论的发现仍是人类在探索宇宙过程中少有的孤胆英雄式的壮举，所得到的结果是前量子时代的物理学皇冠上的明珠。

爱因斯坦的广义相对论之旅开始于一个200多年前牛顿力图回避的问题。引力如何使其影响遍布于无限广阔的空间？遥远的太阳如何影响地球的运动？太阳并未接触地球，那么它是怎样做到这一点的呢？简而言之，引力是如何完成它的工作的？虽然牛顿发现了精确描述引力作用的方程，但他也意识到还留下了一个重要的问题没有解决——引力实际上是如何工作的。在他的《自然哲学之数学原理》一书中，牛顿不怀好意地写道："我把这个问题留给读者思考。"[13]正如我们所看到的，这个问题与19世纪初麦克斯韦与法拉第解决的问题有类似之处，他们用磁场的观点解释了磁铁为什么可以对它实际上并不接触的物体产生力的作用。鉴于上面的启发，你或许会提出一个类似的答案：引力是通过另一种场来发挥作用的，这种场就是引力场。广义上来说，这个答案是正确的。但弄明白这种形式的答案为什么不与狭义相对论矛盾就不是说说那么简单了。

场的办法更加容易。这正是爱因斯坦致力于研究的问题，在接近20年的在黑暗中的探索之后，爱因斯坦取得了这个耀眼的工作成果，爱因斯坦推翻了牛顿有关引力的理论。同样眩目的是，这个问题终于要画上完美的句号了，因为爱因斯坦的关键性突破与牛顿提出的桶的问题有着密切的联系：加速运动的真正本质是什么？

引力和加速等价

在狭义相对论中，爱因斯坦的注意力主要集中在匀速运动的观测者上——观测者感觉不到自己在运动，因此都声称他们是静止的，剩余的世界是运动的。火车上的坏鼠、傻猫和阿布感觉不到任何运动。从他们的角度来看，马丁和月台上的其他人在移动。马丁也感觉他没有运动。对他而言，是火车上的其他人在运动。从他们的角度看，没有谁比谁正确多少。但加速运动是不同的，你可以感觉到它。当汽车向前加速时，你感觉到自己身体向后靠。当火车急转弯时你会感觉到身体被甩向一侧，当电梯向上加速时你会感觉到来自地板的冲击力。

不过，令爱因斯坦吃惊的正是你所感受到的力的平常性。比如说，当你的车急转弯时，为了避免身体被甩向一边，你需要系着安全带，因为这个力是不可避免的，所以只能做点防护措施。没有办法使我们不受这个力的影响，除非我们改变计划不转这个弯了。就是这点使爱因斯坦头脑一震。他意识到引力也具有类似的特点。只要你站在地球上你就会受到地球引力的作用。没有什么方法可以使你避免地球引力的作用。你可以使自己避免受到电磁力和核力的作用，但却没有方法逃脱引力的作用。1907年的一天，爱因斯坦意识到加速与引力之间根本就是相似的。这就是许多科学家们花了毕生精力试图得到的灵光一现，爱因斯坦最终意识到引力和加速运动是同一枚硬币的两面。

正如改变原计划的运动（以避免加速）你就可以避开被车座后背挤压或在车上被甩向一边的感觉，爱因斯坦意识到适当的改变运动，可以避免引力所带来的感觉。这个主意非常简单。为了便于理解，想象一下巴尼，他非常想赢斯普林菲尔德的比赛，所有参加腰带尺寸挑战赛的男性，要在一个月的时间里尽量减肥，最后看看谁减肥减得最多。但是经过两周的液体饮食后（达夫啤酒），他对澡堂的体重秤仍有心理障碍，于是他放弃了所有的希望。紧接着，在澡堂遇到点小麻烦，体重秤粘到了他的脚上，他从澡堂的窗口跳了出去。在他掉下来将要落到邻居的水池时，巴尼回头看了一眼体重秤的读数，他看到了什么？爱因斯坦是第一个意识到巴尼将看到体重秤的读数减为零的

人。由于体重秤和巴尼以同样的速度下落，这样他的脚不再对体重秤施压，所以体重秤的读数为零。在自由落体运动中，巴尼经历了和太空中宇航员同样的经历。

事实上，如果在我们的想象中，巴尼从窗口跳进的是一个大的升降机，其中的空气都被排空，那么在他下降过程中，不仅没有空气阻力，而且由于他身体中每个原子以相同的速率下降，所有平常身体所承受的所有外在压力和牵扯——他的脚冲击着他的踝，他的腿顶着他的臀部，他的胳膊拽着他的肩膀——也都将不存在。[14]下降过程中，闭上眼睛，巴尼将有漂浮在漆黑的太空中的感觉。（我们也可以用无人实验来说明这个问题：用一根绳子拴住两块石头，然后在真空中把它们丢落，绳子始终是松的，就像太空中漂浮的石头一样。）因此，通过改变运动状态——通过完全“屈服于引力”——巴尼能够模拟一个失重环境。（事实上，NASA训练他们的宇航员以适应失重环境与此有异曲同工之妙，NASA让他们的宇航员驾驶经过改装的707飞机，代号叫做“VomitComet”，此飞机的特点是周期性地进入失重状态。）

同理，通过适当的改变运动状态，你就能创造出一种本质上与引力相同的力。想象一下，脚上粘着体重秤的巴尼进入太空船后加入宇航员的失重训练中，体重秤的显示还是零。假如太空船点燃推进器加速，一切就会不一样了。巴尼会感觉到来自太空船地板的压力，就像你站在加速的电梯上感觉到电梯对你的力一样。这时巴尼脚下的体重秤的读数就不再是零了。如果太空船以合适的力度开动推进器，体重秤的读数仍然可以和巴尼在澡堂中看到的一样。通过适当的加速，巴尼能感受到一种与引力难以区别的力。

对于其他种类的加速运动也一样。要是巴尼也加入霍默在太空中的桶中，当桶旋转时，巴尼站在与霍默垂直的方向上——脚和秤都踩着桶内壁——由于他的脚对体重秤施压，秤的读数不再是零了。如果桶以适宜的速率旋转，体重秤的读数与巴尼在澡堂时看到的一样：桶的加速也能模拟地球的引力作用。

以上这些实验和推理使爱因斯坦得出了这样的结论：人们感受到的引力与人们因加速而感受到的力是一样的，它们是等价的。爱因斯坦称其为等价原理。

让我们来看看这意味着什么呢。现在你无时无刻都会感受到引力的影响，如果你站着，你会感觉到地板支撑着你的重量；如果你坐着，你会感觉到其他东西支持着你，甚至你在飞机或火车上读书，你也可能认为你处于静止状态——你没有加速甚至根本没有运动，但根据爱因斯坦的观点，你实际上是在加速。因为你正静静坐着，所以你觉得这种说法听起来有点愚蠢，但不要忘记问问普通问题：加速是以什么为基准来判定的？加速是选谁作参照物的？

有了狭义相对论，爱因斯坦可以宣布绝对时空为狭义相对论提供了基准，但狭义相对论并没有考虑引力的作用。但通过等价原理，爱因斯坦提供了包括引力在内的更加严格的标准。在认知上，这是一场更加激进的变革。既然引力和加速运动是等价的，那么如果你感受到引力的作用，你就一定是在做加速运动。爱因斯坦认为，只有那些根本感觉不到力——包括引力的作用——的观测者才有权利声明他们没有加速。这些不受力的观测者为运动提供了真实的参照系，这种认识，要求我们对思考此类问题的方式做出重大变革。当巴尼从窗口跳到真空的升降机时，通常情况下我们说他向着地球表面加速运动。但这并不是爱因斯坦认同的描述方式。根据爱因斯坦的观点，巴尼并没有加速，他并未感觉到力量，他处于失重状态。他感觉他就像漂浮在漆黑的真空中。他才是所有其他运动的标准。以巴尼为参照物的话，如果你静坐在家中，那么你就在加速。从巴尼的角度——根据爱因斯坦的观点，他的角度才是运动真正的基准——看，当他通过你家窗口自由下落时，你、地球以及我们通常所认为的静止的其他所有物体都在做向上的加速运动。爱因斯坦会认为，是牛顿的头撞上了苹果，而不是苹果撞上了牛顿。

很明显，这是一种全然不同的考虑运动的方式。它所依靠的是这

样简单的一个认识：只有当你对抗引力时你才能感觉到引力的作用。相反，当你完全无须抗拒引力时你就感觉不到它的存在了。假设你没有受到任何其他影响（比如空气阻力），当你屈服于引力使得自己自由下落时，你会感觉自己就像自由漂浮在真空中——这种状态，不用怀疑，我们认为是没有加速。

总之，只有那些自由漂浮的个人，不管他们是在太空深处，还是在撞向地球表面的过程中，才有权利声明他们并未加速。如果你经过一个这样的观测者，你们之间有相对加速，那么根据爱因斯坦的观点，你就是在加速。

事实上，坏鼠、傻猫、阿布和马丁中没有一个人有资格说他们在决斗中是静止的，因为他们都感觉到引力竖直向下的作用。但这不会影响我们前面的讨论，因为以前我们关心的只是水平运动，竖直方向上的运动不会影响水平运动。但作为一条重要的原理，爱因斯坦发现的引力和加速运动之间的联系意味着，再说一遍，只有那些感觉不到力的观测者才能真正地说他们处在静止状态。

爱因斯坦已经发现了引力和加速运动之间的联系，接着开始准备结果牛顿遗留的挑战，寻求引力是如何发挥其影响的合理解释。

蜷曲、弯曲与引力

通过狭义相对论，爱因斯坦指出每个观测者把时空切成平行的片，这些片可以看成是一系列连续时间段的空间；相对于彼此匀速运动的观测者将会从不同的角度切割时空。如果这样一个观测者开始加速，你可能会认为他每一时刻速度的改变或运动方向的改变将会导致他的切片的方向和角度不停地改变。简单地说，就是这样。爱因斯坦（利用卡尔·弗雷德里希·高斯，格奥尔格·伯恩哈德·黎曼以及19世纪其他数学家建立起来的几何学知识）发展了这个观点——断断续续的——指出从不同的角度切割时空将会形成弯曲的片，但合起来却像银

盘里的调羹一样完美，如图3.8所示。加速运动的观测者将切出弯曲的时空片。

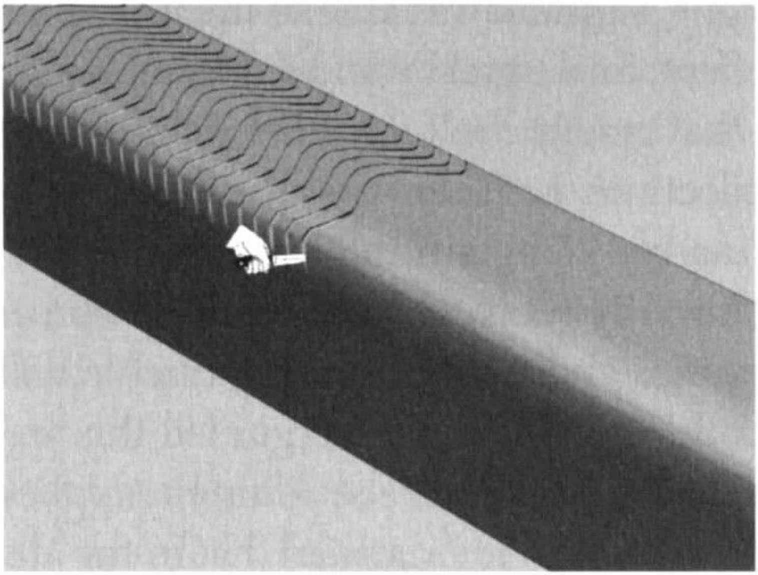


图3.8　根据广义相对论，不仅可以从不同角度（由处于相对运动中的观察者来操作）把时空面包切成片（每一片都代表某一时刻的空间），而且切片本身可以因物质或能量的出现而发生弯曲。

有了这样的观点，爱因斯坦就能使等价原理产生深刻的影响。既然引力和加速是等价的，爱因斯坦领会到，引力本身不是别的，正是时空结构中的蜷曲和弯曲。让我们来看看这究竟是什么意思。

如果你沿着光滑的木制地板滚一颗弹球，它的轨迹将是一条直线。但如果最近发洪水了，洪水退后，地板上全是坑坑洼洼，那么滚动的弹球就不再沿着原来的路径了。它将受路的引导，受地表坑坑洼洼的影响。爱因斯坦把这个简单的观点用在宇宙结构上。他假想了一种情景，没有物质和能量——没有太阳，没有地球，也没有其他星球——的时空，就像光滑的木制地板一样没有坑坑洼洼。它是平的。如图3.9（a）所示，其中，我们关注的是一个空间片。

当然，因为空间是三维的，所以图3.9（b）更加精确些，但是把图画成二维的更便于理解，因此我们继续使用它。爱因斯坦认为物质和能量的存在对空间的影响就像洪水会对地面的影响一样。物质和能量，比如说太阳，会使空间（和时空[9]）像图3.10（a）和3.10（b）所描述的那样发生弯曲，就像在崎岖的地板上滚动的弹球是沿着崎岖的路径滚动一样。爱因斯坦指出，在蜷曲空间穿行的任意事物——就像地球在太阳周围运行一样——将会沿着弯曲的轨迹运行，就像图3.11（a）和3.11（b）所描述的那样。

物质和能量在时空结构中留下了深沟众壑，物体就好像被无形的手引导一样，沿着时空结构中的沟壑运动。根据爱因斯坦的观点，这就是引力施加影响的方式。同样的观点更适用于家中。

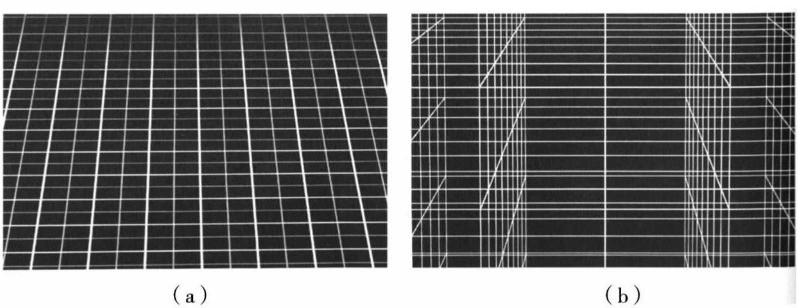



图3.9　（a）：平坦的空间（二维）。（b）：平坦的空间（三维）。

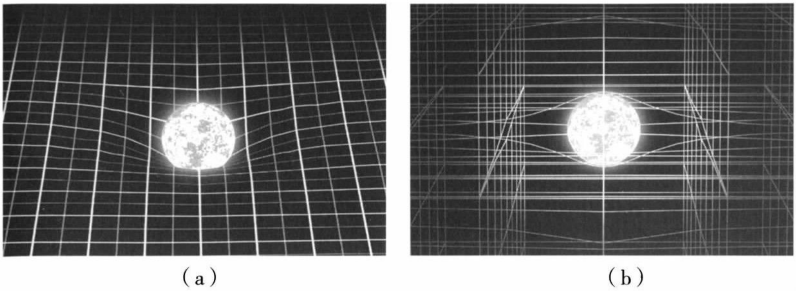
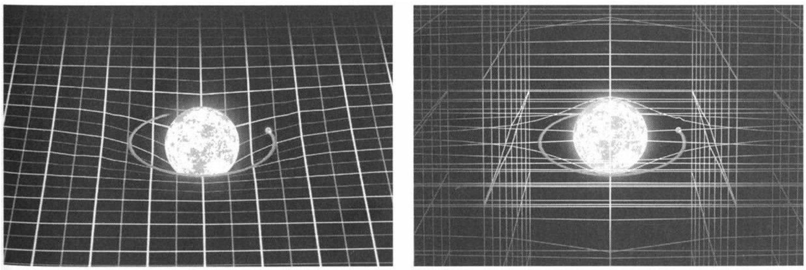


图3.10　（a）：太阳使空间发生弯曲（二维）。（b）：太阳使空间发生弯曲（三维）。



（a）　（b）

图3.11　地球绕着太阳运行，因为它沿着因太阳出现而造成的时空结构的弯曲而运行。（a）：二维版本。（b）：三维版本。

现在，你的身体有沿着由于地球的存在而产成的时空结构的凹痕滑下的倾向，但是你的下滑趋势被你正坐在或站在上面的地球表面阻挡。你生命中的每一刻都能感受到的向上的推力——来自地板，房间的地面，角落里的便椅或是你的双人床——都在阻止你沿着时空的峡谷滑下来。与之相反的是，如果你将自己高高置于跳水板上，让你自己完全屈服于引力，你就会随着时空瀑布自由下落。

图3.9、图3.10、图3.11几个示意图描绘出了爱因斯坦10年奋斗的成果。爱因斯坦这些年来的大部分工作旨在精确定出一定数量的物质和能量所造成的时空弯曲的形状和大小。爱因斯坦发现数学成果为这些图奠定了基础，它们在爱因斯坦场方程中有具体的表达。正如其名字所启示的一样，爱因斯坦把时空的弯曲看做引力场的表现——几何表现。通过在几何上思考这个问题，爱因斯坦发现方程在解决有关引力问题中所起的作用就像麦克斯韦的方程在电磁学中所起的作用一样。[16]爱因斯坦和其他科学家们通过这些方程预言了各个星球的运行轨迹，甚至还有遥远的星球发出的光穿过弯曲时空的轨迹。这些预言不仅有很高的精确度，而且可以与牛顿理论在预测方面的成就相媲美，爱因斯坦的理论终于以更高的准确度与实际情况相符。

同等重要的是，由于广义相对论明确了引力工作的细节，因而为确定引力的传播速度提供了重要的数学框架。引力传播速度问题变成了空间的形状改变得有多快的问题。即，空间蜷曲和起涟漪——就像是石子被扔进水池所引起的涟漪——在空间中的传播速度究竟是多少？爱因斯坦能把这个计算出来，而且答案令人非常满意。他发现空间的蜷曲和涟漪——即引力——从一个地方到另一个地方并不是牛顿式的引力计算的那种瞬时传播。相反，它们以光速传播，与狭义相对论设置的速度极限完全一致。如果外星人把月球弄没，潮水会在一秒半后以我们看到月球消失的速度退去。牛顿理论失败的地方正是爱因斯坦的广义相对论发挥优势的地方。

广义相对论和桶

广义相对论不仅构建出数学上优美，概念上有力，并且第一次实现自洽的引力理论，而且也彻底重塑了我们的空间观和时间观。在牛顿的理念和狭义相对论中，空间和时间为宇宙事件提供了一个不可变更的舞台。在狭义相对论中，将宇宙分割为连续的时间片——每一片都表示某个时刻的空间——的方式可以有很多种，尽管在牛顿时代的人眼中这有点不可理解，但时间和空间还是不会对宇宙中发生的事件做出回应。时空——我们一直用面包条来指代的时空——被看做是一成不变的。但在广义相对论中，一切都不一样了。时间和空间成了宇宙演化的参与者，它们变得生动起来。物质使空间发生蜷曲，而蜷曲又使物质移动，从而进一步使空间弯曲。广义相对论为空间、时间、物质和能量在宇宙中的舞蹈提供了广阔的舞台。

这是一个令人惊奇的进展。但现在我们回到我们的中心主题：桶怎么样了？广义相对论如爱因斯坦所希望的那样为马赫的相对主义观点提供了物理基础吗？

多年来，这个问题引起了许多争议。最初，爱因斯坦认为广义相对论完全包括了马赫的观点，他认为马赫的观点非常重要，并将其命

名为马赫原理。事实上，1913年的时候，当爱因斯坦疯狂地工作以期完成广义相对论的最后一块拼图时，他给马赫写了一封热情洋溢的信，在信中他描述了广义相对论究竟是怎样肯定了马赫对牛顿水桶实验的分析。[17]1918年，爱因斯坦写了一篇文章，列举了广义相对论背后的三个重要观点，他所列举的第三个即是马赫原理。但是广义相对论十分奥妙，对其性质的研究花了包括爱因斯坦在内的科学家的多年时光。当这些困难方面渐渐理清之际，爱因斯坦发现很难把马赫原理纳入到广义相对论中。慢慢地，爱因斯坦从马赫原理中醒悟过来，在他生命的最后几年，他逐渐断绝了与马赫原理的关系。[18]

又经过了半个世纪的研究之后，我们事后诸葛亮，重新审视一遍广义相对论对马赫原理的符合程度。尽管还有一些争论，但我认为最准确的说法是，在某些方面，广义相对论显然有一些马赫观点的意味，但是它并不完全符合马赫提倡的相对论观点。下面就是我要讲的意思。

马赫认为，[19]当旋转的水面开始凹陷时，或者当你感觉到胳膊向外伸展时，或者当两块石头之间的绳子拉紧时，这些都与假想的——按他的观点，完全是误导——绝对空间（或从我们现在的观点来看的绝对时空）概念毫无关系。相反，他提出加速运动很明显是相对于整个宇宙中的所有物质而言的。没有物质，就没有加速运动的概念，也没有所列举的一系列物理效应（凹陷的水面，张开的胳膊，拉紧的绳子）。

广义相对论又是怎么说的呢？

根据广义相对论，所有运动特别是加速运动的基准是自由落体的观测者——完全屈服于引力且没有感受到任何其他力作用的观测者。现在，关键的一点是，作用于自由落体状态的观测者的引力来源于整个宇宙中传播的所有物质（和能量）。地球，月球，遥远的行星，恒星，大气层，类星体和其他星系都对你所在的位置产生引力场（用几

何语言来说是时空的曲率）。体积较大、距离较近的物体产生的引力影响更大些，但是你所感受到的引力场代表着其他所有物体作用的综合影响。[20]你完全屈服于引力做自由落体运动——你成为判断其他物体是否加速的标准——时走过的路径将会被宇宙中所有物质影响，既包括天空的恒星又包括隔壁的房间。因此，在广义相对论中，当我们说一个物体加速时，就意味着这个物体相对于宇宙中所有物质所决定的基准加速运动。这个结论正是马赫所倡导的。因此，从这个意义上来讲，广义相对论确实在一定程度上包含了马赫的想法。

然而，广义相对论并没有肯定马赫所有的推理，就像我们考虑——再一次的——在没有其他物质的真空宇宙中旋转的桶时直接所看到的那样。在一个空的、不会变化的宇宙中——没有恒星，没有行星，没有任何东西——没有引力。[21]没有引力，时空就不会弯曲——这时它所呈现的是3.9（b）所示的简单的未弯曲的形状——这就意味着我们回到了狭义相对论所描述的更为简单一点的情况中（记住，在提出狭义相对论时爱因斯坦忽略了引力。广义相对论通过包含引力弥补了这个缺陷。但是，当宇宙中什么都没有且不变化的时候，不存在引力，广义相对论就简化成了狭义相对论）。如果我们把桶引进到真空宇宙中，由于它的质量太小了，所以它的存在根本不会影响空间的形状。因此，我们以前用狭义相对论对桶进行的有关讨论同样也适用于广义相对论。这与马赫所预言的不一样，广义相对论得出了和狭义相对论一样的结果，并声明即使在没有其他物质的真空宇宙里，你依然可以感觉到旋转的桶的内壁的压力；你旋转时，你的胳膊会感觉到向外拉的张力；系在两个旋转的石头之间的绳子仍然可以被拉直。我们所得到的结论是，即使在广义相对论中，空的时空也可以为加速运动提供一个基准。

因此，虽然广义相对论采纳了马赫想法的一些元素，但它并未完全从属于马赫所提倡的相对论观点。[22]马赫原理是一种富于启发性的思想，它的确为革命性的发现提供了灵感，即使这最终的发现并未将

激发其灵感的观点纳入体系中。

**3000年的时空**

旋转的桶已经讨论了很长时间。从牛顿的绝对空间和绝对时间，到莱布尼茨和后来马赫的相对论观点；再到爱因斯坦在狭义相对论中认识到的：空间和时间是相对的，但它们组合起来却是绝对时空；再到他的下一个发现，在广义相对论中认识到的：时空是动态宇宙中的一个参与者，在这个过程中，那只旋转的桶总是在那儿。在我们内心深处，旋转的桶提供了一个简单、静态的检验方式，以探明不可见的，抽象的，不可触摸的空间（从广义上来说是时空）是否足以为运动提供最终的参照物。裁判？虽然这个问题还在争论之中，但正如我们所看到的，对爱因斯坦和他的广义相对论最直接的解读就是，时空可以提供这样一种基准：时空的确是具体的。[23]

注意，这个结论也是更广意义上的相对主义支持者欢呼庆祝的原因。从牛顿和后来的狭义相对论的角度来看，空间和后来时空被看做是可以为定义加速运动提供基准的实体。因此，根据这些观点，空间和时空是绝对不可变的，加速的概念是绝对的。在广义相对论中，时空的特点就完全不同了。在广义相对论中，空间和时间是动态的：它们是可变的；它们随物质和能量的出现而不同；它们并不是绝对的。时空，特别是它蜷曲和弯曲的方式正是引力场的体现。因此，在广义相对论中，相对于时空的加速运动远非以前的理论提出的绝对的、与相对无关的概念。相反，爱因斯坦，在其逝世前一些年[24]曾意味深长地说，相对于广义相对论中的时空而言，加速运动是相对的。这里所说的加速并不是相对于石头、恒星一类的实物而言，而是相对于某种真实的、可触摸的、可变的事物而言的，这种事物就是，场——引力场。[10]从这种意义上来讲，时空——通过引力得以彰显——在广义相对论中是如此真实以至于它所提供的基准是许多相对主义者可以舒心接受的。

当我们真正开始理解空间、时间和时空究竟是什么时，我们就本章中所讨论的问题而展开的争论无疑将会继续下去。随着量子力学的发展，问题变得更加厚重。当量子力学登上历史舞台时，真空和虚无的概念呈现出全新的意义。事实上，自打爱因斯坦在1905年废除了光以太，关于空间充满不可见的物质的观点一直试图重新回来。我们在后面章节中将会提到，现代物理学中的关键发展一直在以各种各样的形式重塑以太类物质，不过这些新的以太类概念并不像原始以太概念那样要为运动提供绝对的标准，但是所有的这些以太概念全部会对全空的时空这样的幼稚观点发起挑战。而且，正如我们将要看到的，空间在经典宇宙中所起的最基本作用——比如作为隔离物体的介质，使我们可以明确说明物体彼此区别、彼此独立的无所不入的事物——将遭遇惊人的量子概念的全面挑战。

## 第4章　纠缠着的空间

在量子宇宙中，分隔是什么意思

接受狭义相对论和广义相对论就意味着放弃牛顿的绝对空间和绝对时间观念。尽管并不容易，但你可以训练自己的思维以做到这一点。当你走来走去时，想象着不随你运动的事物所感受到的此刻与你所感受的此刻是不一样的。当你在高速路开车时，想象着你的手表运转的速率与在家中时是不一样的。当你从山顶上四下张望时，想象着由于时空的弯曲，相比于那些在低处受到更强引力的物体，时间于你而言变得更快了。我之所以说“想象”，是因为在类似的普通情况下，相对论的作用如此之小以至于它们完全不会被注意到。日常经验无法展现宇宙究竟是怎样运行的，而这也就是为什么在爱因斯坦提出相对论之后的一百多年中，没有一个人，即便是专业物理学家，在他的骨骼中也没有感受到相对论的存在。这并不奇怪：人类只有在巨大压力下才有可能得到来自于牢靠掌握相对论的生存优势。而在我们日常生活中所遇到的低速、适度重力的情况下，牛顿关于绝对空间和绝对时间的错误概念运作得非常好，因此，我们的感觉在没有进化压力的情

况下无法发展出能感受到相对效应的敏锐性。所以，我们需要努力运用智力以弥补感官上的不足，才能达到真正的觉醒与理解。

相对论代表了对宇宙传统观念的里程碑式的突破，而在1900～1930年，另一场革命使物理学陷入了混乱之中。它开始于20世纪之初的两篇关于辐射性质的论文，一篇是马克斯·普朗克写的，另一篇是爱因斯坦写的。这两篇文章，在30年的热烈研究后，终于导致了量子力学的诞生。相对论的效应只在极端速度或引力条件下才有意义，和它一样，新物理——量子力学——也只在另一种极端条件下才能充分显现，这种极端条件就是超微观世界。但在相对论带来的剧变和量子力学带来的剧变之间有着明显的区别。相对论的神秘性源于我们每个人的时空体验不同于其他人的时空体验，它的神秘性来自于比较。我们不得不承认我们对于实在性的看法只能算是满足时空要求的众多可能性——事实上是无穷多的可能性——中的一种。

量子力学则不同。它的神秘性显然是无须比较的。很难将你的思维训练成拥有量子力学式的直觉，因为量子力学打破了我们个人对于实在性的概念。

量子眼中的世界

每个时代都有自己关于宇宙孕育形成的观点。根据古印度创世神话，神肢解了原始巨人普鲁萨，将他的头颅化作天空，双足化作大地，呼吸变作风，于是宇宙就这样形成了。对于亚里士多德而言，宇宙是55个同心水晶球的集合体，最外层的是天堂，接着是行星，地球和它的自然环境，最后是七重地狱。[1]牛顿和他有关运动的精确的、确定性的数学体系，再次改变了对宇宙的描述。宇宙就像是一个巨大宏伟的时钟：上紧发条设定好初始状态后，宇宙就会以高度的规律性和可预见性滴滴答答地从这一刻走到下一刻。

狭义相对论和广义相对论指出了时钟这个比喻的重要细节：根本

不存在一个优先的、普遍通用的时钟；关于什么构成了时刻，什么构成了现在没有普遍共识。即便这样，你依然可用一个时钟式的宇宙进化故事。这个时钟是你自己的时钟，这个故事是你自己的故事，但宇宙却会像牛顿体系中的那样呈现出规律性和可预见性。如果你通过某种方法知道了宇宙现在的状态——如果你知道了每个粒子在哪，朝向哪个方向，以多快的速度运动——那么，牛顿和爱因斯坦都将同意，从原则上讲，你可以运用物理定律预测未来任意时刻宇宙中将发生的一切，也可以描述过去任意时刻宇宙所曾发生过的一切。[2]

量子力学破坏了这个传统。我们永远也不可能知道单独一个粒子的精确位置和速度。我们甚至不可能完全确定地预测哪怕最简单实验的结果，就更不用说整个宇宙的演化了。量子力学告诉我们，我们所能做的事情，不过是预测实验得出或这或那的结果的概率。随着量子力学经受住了几十年来各种精确到难以想象的地步的实验检验，牛顿式的宇宙时钟，甚至是爱因斯坦的升级版，都变成了站不住脚的比喻；世界显然不是这样运转的。

但过去的突破还是不完全的。尽管牛顿和爱因斯坦的理论在空间和时间的性质上，观点完全不同，但两者却在某些基本方面，某些不证自明的真理上保持一致。如果两个物体之间有空间——比如天空中有两只鸟，一只在你的左边，而另一只在你的右边——那么我们就可以认为这两个物体是独立的。我们把它们看做是可以分开的、有所区别的实体。空间，不管从根本上讲它是什么，提供了分割并区分物体的介质。空间干的就是这些事情。位于空间中不同位置的物体就是不同的物体。而且，要想使一个物体影响另一个物体，就必然会以某种方式与隔离它们的空间打交道。一只鸟飞向另一只鸟，穿过它们之间的空间，才可以啄到或靠近另一只鸟。如果一个人想影响另一个人，他可以使用弹弓，使一个小石子穿越他们之间的空间；或者他可以大声喊叫，使活跃的空气分子形成多米诺效应，分子一个碰一个，直到传至接收者的鼓膜。一个人想对另一个人施以影响还有更复杂的方

式，发射激光，使电磁波——一束光——穿越他们之间的空间；或者，更野心勃勃的人（就像在上一章中提到的外星恶作剧者），可以摇动或移动巨大的物体（如月球），将引力扰动从一个地方传到另一个地方。可以肯定的是，如果我们想从某个地方影响另一个地方的人，不管我们怎么做，这个过程总是会涉及将某人或某物从一个地方传送到另一个地方，并且只有当某人或某物到达另一个地方时，影响力才能发挥出来。

物理学家们把宇宙的这种特性称为定域性，以强调你只能对你附近的物体产生直接影响。魔法就违反定域性，因为使用魔法就可以通过在此地做某事影响彼地的某物，而不需要任何东西从一个地方到另一个地方，但常识却告诉我们那是不可能的，可重复的实验会证实定域性。[3]而大部分的实验的确也做到了。

但是20世纪末的一些实验却表明，我们在此地做的某件事情（如测量一个粒子的某种特性）可以巧妙地与彼地发生的某件事情（如测量某个相距较远的粒子的某种特性所得到的结果）发生联系，而无须把一个物体从某地移到另一地。虽然直观上有点令人困惑，但这种现象完全符合量子力学定律；早在相应的技术发展到可以做该实验时，量子力学就已经预言了这种实验结果，而现在实验证明，该预言的确是正确的。这听起来有点像魔法；爱因斯坦，作为最早认识到——并尖刻地批判——量子力学这种可能的特征的物理学家之一，将其称之为“鬼魂一般”。但是我们将会看到，这些实验所证实的长距离间的联系是相当敏锐的；而且，在精确的意义上，从根本上超出了我们能力的控制范围。

不管怎样，来自理论和实验的这些结果，强有力的支持这样的结论：宇宙允许非定域关联的存在。[4]发生在此地的某事可以与发生在彼地的事情有所关联，即便没有任何东西被从一个地方移动到另一个地方——即便没有足够的时间让任何东西，即使是光，在两地之间传播——也没有关系。这就意味着空间不可能像人们曾经认为的那样；

实际上，两个物体间的空间，不论有多大，都不能确保这两个物体是分离的，因为量子力学允许它们之间存在纠缠，或者说某种关联。一个粒子，比如构成你或我的无数粒子中的一个，能运动但却无法隐藏起来。根据量子理论以及证实了很多预言的实验，即便两个粒子分别处于宇宙的两端，它们之间的量子关联依然存在。从它们之间相互关联的角度来看，尽管它们之间隔着几万亿英里的空间，但看起来就像其中一个在另一个正上方一样。

对我们实在性概念的攻击，大都来自于现代物理学；在后面的章节里我们还将会遇到很多。但就这些已经被实验证实的结果来看，没有什么比认识到宇宙不具有定域性更令人难以置信的了。

红和蓝

为了体验一下量子力学的这种非定域性，想象一下史考莉探员[11]，她很长时间没有休过假了，现在返回她在普罗旺斯[12]的别墅度假。还没等她打开包裹，电话铃声就响了。是在美国的探员穆德打来的。

“你拿到那个包裹了吗——用红蓝纸包装的那个？”

史考莉走到门口，她把所有的邮件都堆在那里，看到了那个包裹。“穆德，别这样，我跑这么老远来到爱克斯不是来处理另外一堆文件的。”

“不，不，你搞错了，这个包裹不是我寄给你的。我也收到了一个，里面装着些很小的不透光的钛盒，从1标到1000，有封信说你将会收到同样的一封信。”

“是的，看到了，怎么回事。”史考莉慢慢回答道，担心这些不透光的小钛盒会破坏她的假期。

“这个嘛，”穆德继续说，“信上说每个小钛盒都装着一个不同的小球，一打开上面的门这些小球就会发红光或蓝光。”

“穆德，是要给我什么惊喜吗？”

“恐怕不是这样。听着，信上说在任何一个盒子打开之前，这个小球都有可能在发红光或蓝光，而在门打开的一瞬间，到底是什么颜色的光是随机决定的。但奇怪的是，信上说虽然你的盒子和我的盒子完全一样——盒子里的球也都是随机发出红光或蓝光——但我们俩的盒子却总是一唱一和的。信上说它们之间有种神秘的联系，我要是打开1号盒子时发现有蓝光的话，你打开你的1号盒子时也会发现蓝光；如果我打开我的2号盒子时发现有红光的话，你打开你的2号盒子时也会发现红光，如此种种。”

“穆德，我真的筋疲力尽了；等我回去再玩这些小把戏吧。”

“史考莉，请别这样。我知道你在休假，但是我们不能不管这些事情啊。我们只需要几分钟看看这是否是真的。”

虽然不情愿，但史考莉认识到抗议是无效的，于是史考莉走过去打开她的盒子。通过对比每一个盒子中所发出的光的颜色，史考莉和穆德发现确实如信上所说的一样。盒子中的小球有时发红光，有时发蓝光，但是打开相同号码的盒子，史考莉和穆德发现盒子中发出的总是相同的光。穆德对此现象感到十分兴奋，但史考莉却没有太大兴趣。

“穆德，”史考莉在电话里严肃地说，“看来你也需要一个假期了。这太傻了。显然，盒子的门被打开时，每个盒子里的小球都是按编好的程序发红光或蓝光。他们送给我们的盒子编有相同的程序，因此我们打开相同号码的盒子它们都会发相同颜色的光。”

“但是，史考莉，信上说当盒上的小门打开时，小球随机发出红

光或蓝光，而不是按预先编好的程序发某一种光。”

“穆德，”史考莉说，“我的解释听起来更合理些，并与事实相符。你还想要怎样？看这儿，信的最底部，这是最令人可笑的。‘外星人’的小印刷字体告诉我们，不仅打开盒子的小门会造成里面的区域发光，其他用来弄清楚盒子是怎样运行的方式——比如，如果我们在门打开之前，先检查盒子这些区域的颜色组成或是化学物质的构成——也同样会使它发光。换句话说，我们无法分析他们所假定的在选择红色或蓝色上的随机性，因为任何这样的尝试都将会影响我们试图得出的实验结果。这就像我告诉你我是金发，但是当你或其他人又或什么东西看我的头发或分析它时，我就变成了红头发。你怎样证明我是错误的？给你这些盒子的人是非常聪明的——他们已经把事情都安排好了，它们的诡计不会被识破。现在，赶紧去玩弄你的那些小盒子吧，而我想要静静地待会儿。”

看来史考莉已经站在了科学一边。但是，实际是这样的。80年来，量子工程师们——科学家们，而不是外星人——一直宣称宇宙的运行方式正如上文中的信上所言。问题是，现在有强有力的科学证据表明，穆德这边——而不是史考莉——才是数据所支持的一边。举个例子来说，根据量子力学，一个粒子可以处于一种特性和另一种特性间的边缘地带——就像在盒子的小门打开之前，一个“外星”小球可以徘徊于发出红光或发出蓝光之间——只有当人们关注（即测量）粒子时，它才会随机地选择其中的一种。好像这也没什么奇怪的，量子力学也是预言粒子之间存在联系，就像前面提到的那些外星小球之间的联系。两个粒子因量子效应而产生的纠结如此强烈，以至于它们在随机选取某种特性上也是相互关联的：就像每个外星小球随机选择发红光或蓝光，但相同号码的盒子里的小球所选择的颜色却是彼此关联的（都发红光或发蓝光）一样，两个粒子，即便在空间上相隔很远，它们随机选择的特性也会存在类似的某种完美的关联性。简单地说，量子力学告诉我们，即使是相距很远的两个粒子，只要其中一个做了什

么，另一个也会做相应的事情。

举一个具体的粒子，假如你戴了一副太阳镜，量子力学就会告诉我们，一个特殊的光子——比如从湖面或沥青路面反射到你身上的那种——将有50%的概率穿过你佩戴的偏光眼镜，从而使进入你眼睛中的光强减弱：当光子碰到你的镜片时，它会“随机”选择反射回来或穿过眼镜。令人惊奇的是，这样的光子可以有一个在几英里以外向相反方向传播的伙伴，另一个光子也会以50%的概率通过另一个偏振太阳镜，就像它的伙伴那样。即便每一个结果都是随机决定的，即便它们在空间上相距很远，但只要一个粒子能通过偏振镜，另一个光子也将通过。这就是量子力学预言的所谓非定域性。

爱因斯坦本人从来都不是量子力学的推崇者，他不愿意相信宇宙会按这样诡异的原则运行。他支持更为传统的解释，在这样的解释中，不存在粒子会在被测量时才随机选择某种特性或结果这样的概念。与之相反，爱因斯坦认为，即便人们观测到两个距离很远的粒子拥有相同的特性，也并不能证明存在着与粒子属性有瞬时联系的神秘的量子关联。而且，就像史考莉认为小球并不是随机选择发出红光或蓝光，而是被提前设计好了当被观测时发哪一种光，爱因斯坦认为，粒子并没有随机选择哪一种特性，而是类似的被“设计好了”当被观测时拥有哪一种特殊的、具体的性质。爱因斯坦说，相距很远的粒子在属性上的关联，证明了光子在被发射时具有相同的性质，而不是遵循某种诡异的长程量子纠缠。

将近50年过去了，关于这个问题的谁是谁非——爱因斯坦对还是量子力学的支持者们对——仍未解决，因为正如我们所看到的，这个争论就像是史考莉和穆德之间的争论：任何试图否定奇怪的量子力学关联且同时保有更为传统的爱因斯坦观点的努力都将是徒劳的——实验本身必然会污染试图得到的实验结果。所有这一切在20世纪60年代都变了。凭借着令人叹服的洞察力，爱尔兰物理学家约翰·贝尔指出这个问题可以用实验来解决，并且，人们的确在20世纪80年代完成了这

个实验。从数据得出的直接结果来看，是爱因斯坦错了，距离很远的物体之间的确可以有奇妙诡异的量子关联存在。[5]

这个结论背后的论证如此繁复，以至于物理学家们用了30多年的时间才完全弄清楚一切。但是，在搞清楚了量子力学的本质特征后，我们将看到，这一论证的核心思想不会比一道脑筋急转弯还难。

产生波

取一小张黑色的35毫米胶片，沿着两条非常紧密的细线刮去感光乳剂，然后用激光去照射它，你将看到光是一种波的直接证据。如果你没这样做过，很值得一试（你可以用许多东西来取代胶片，比如咖啡机中的金属丝网）。当激光通过胶片的细缝碰到探测屏时，就会使探测屏上出现如图4.1所示的明暗相间的条带，解释这种图形的出现需要知道一些波的基本特性。水波是最常见的波，因此我们就先以平静的湖面上出现的水波为例，解释一下波的重要性质，然后我们再将讨论推广到光波的解释上。

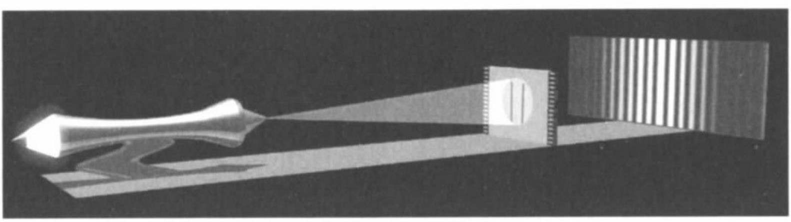


图4.1　激光通过黑色底片上的两条细缝在探测屏上产生干涉图样，表明光是一种波。

水波会在平静的湖面上激起涟漪，使得有些地方隆起，有些地方凹下去。波最高的部分称为波峰，最低的部分称为波谷。典型的波具有周期性，波峰后是波谷，波谷后是波峰，依此类推。如果两列波相遇——举个例子来说，假如你我在湖面上相近的位置各扔下一颗小石子，那就会产生相向的波——当它们相遇时将会产生如图4.2（a）所示的干涉效应。当一列波的波峰和另一列波的波峰相遇时，干涉波的

振幅将是前面两列波的波峰的叠加。类似的，当波谷与波谷相遇时，水面将会由于两个波谷叠加而凹陷得更深。而最为重要的情形是：当一列波的波峰与另一列波的波谷相遇时，它们将相互抵消，波峰试图将水往上提，而波谷试图将水往下拉。如果一列波的波峰高度与另一列波的波谷深度相等，它们将完全抵消，这样的话该位置的水根本就不会上下移动。

相同的原理也可以用来解释图4.1中一束光通过两条细缝时产生的图样。光是一种电磁波，当它通过两个细缝时会分裂成两列波，这两列波会在探测屏上相遇。就像上文所讨论过的水波一样，这两列光波发生了干涉。它们在探测屏上不同位置相遇时的情形也不一样，有时波峰和波峰相遇，于是探测屏上的相应位置就会变得更亮；有时是波谷和波谷相遇，那么探测屏上的相应位置也会变得更亮；但当一列波的波峰遇到另一列波的波谷时，它们就会相互抵消，于是探测屏上的相应位置就会变得黯淡。如图4.2（b）所示。

用数学仔细的分析波的运动，包括两列波的波峰和波谷之间的各个不同位置之间部分抵消的情况，我们会发现，计算所得到的明暗分布正如图4.1所示的那样。因此明暗相间的条带的出现说明光是一种波，自从牛顿认为光不是一种波而是由一束粒子组成的（稍后还要就此进一步讨论）以来，这个问题已经争论很久了。我还要指出，这里的分析也同样适用于任何一种波（光波，水波，声波，总之你能叫得上名的任何一种波），而且，干涉图样为我们提供了波的证据：当波被迫通过两条适当大小的细缝时（大小由波峰和波谷之间的距离所决定），就会产生图4.1中所示的加强图样（亮的条带代表着高强度，黑色条带代表着低强度），于是我们就知道它的确是一种波。

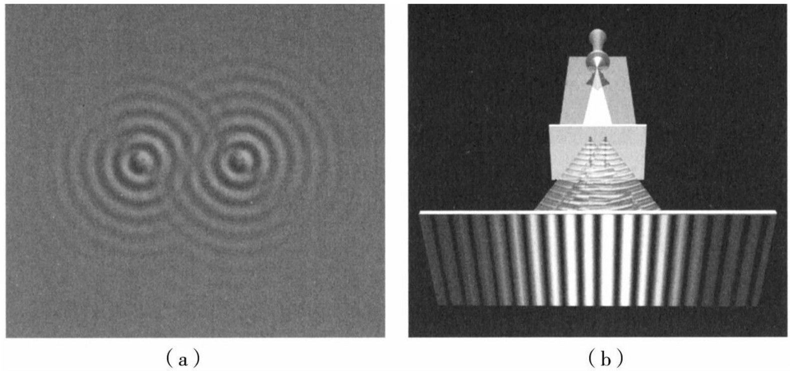


图4.2　（a）：重叠的水波产生了干涉图样。（b）：重叠的光波产生了干涉图样。

1927年，克林顿·戴维森和莱斯特·戈默用一束电子——与波没有任何明显联系的微粒实体——穿过一片镍晶体；细节我们并不关心，重要的是这个实验等价于用一束电子撞击两条细缝之间的障碍物。在该实验中，电子穿过细缝射到磷屏上，它们打在磷屏上的位置将会以小亮点的方式记录下来（我们平常所看的电视机的屏幕也是由类似的小亮点组成的），实验结果令人非常惊奇。想象一下，把电子看做小子弹或小球，我们自然就会想到它们撞击磷屏的位置会与这两条细缝平行一致，如图4.3（a）所示。但那并不是戴维森和戈默得到的结果。他们的实验产生了如图4.3（b）所示的结果：电子撞击磷屏后，呈现出了波所特有的干涉图样。戴维森和戈默发现了干涉图样这个证据。他们的结果出人意料的表明，电子束一定是某种波。

现在，你也许觉得这个结果并不是特别令人惊讶。水是由$H_2O$分子组成的，当许多分子以一定模式移动时水波就会出现。一组$H_2O$分子在某些位置上升，在邻近的位置，另一组$H_2O$分子却在下降。或许图4.3所示的数据只是表示电子，会像$H_2O$分子一样，有时以一定模式移动，其整体的宏观运动形成了类似于波的干涉图形。乍一看，这似乎是个合理的解释，但事实上，真实的情况更加令人出乎意料。

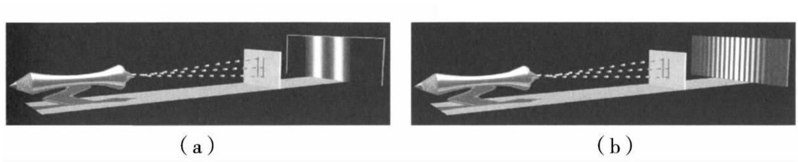


图4.3　（a）：经典物理学预言电子通过障碍物的两条裂缝后将在探测屏上产生两条明亮的条带。（b）：量子物理预言并且也得到了实验的证实，电子将会产生干涉图样，这表明它们具有波的特性。

我们来想象一下，如图4.3那样，从电子枪持续不断地射出电子束。但我们能调节枪使它每秒钟射出的电子越来越少；事实上，我们可以把它调到很低，比如说，使其每10秒射出一个电子。有足够耐心的话，我们可以用很长时间来做这个实验，记录穿过细缝的每个电子撞击探测屏的位置。图4.4（a）~（c）分别显示了在1小时、半天以及一天内所积累起来的数据的结果。20世纪20年代，类似于这样的图像构建起了物理学的基础。我们可以看出，即使单独的电子，分别的、一个接一个的撞击在探测屏上，也会形成波的干涉图样。

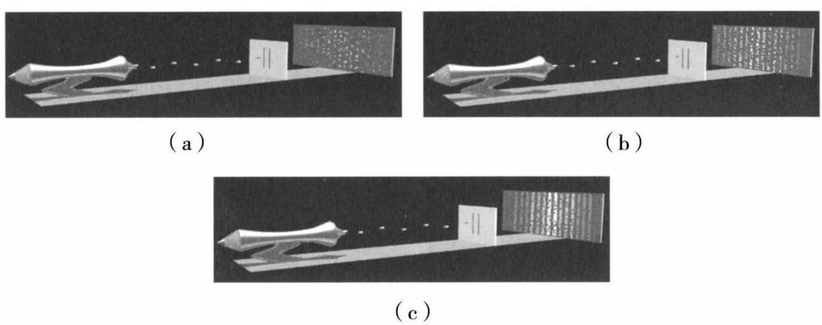


图4.4　发射出去的一个个电子通过裂缝后在探测屏上形成了干涉图样。图（a）~（C）描述了电子形成干涉图样的过程。

这就好像在说一个单独的水分子也可以形成水波。但那怎么可能呢？波的运动看起来是某种集群性质，当应用于单独的微粒时就没有意义了。比如说，球场的看台上，时不时地就会有个别的观众站起来或坐下，但这种起来坐下无法形成人浪。而且，波的干涉似乎需要两

列来自不同地点的波相遇。因此波的干涉怎么可能只与单独的微粒组分有关？但不管怎么说，正如图4.4中干涉图样所证明的那样，即使单个电子是微小的物质粒子，它也可以具有波的某些特点。

概率和物理学定律

假如一个单独的电子也是波，那么是什么在波动呢？欧文·薛定谔迈出了第一步：或许构成电子的物质在空间杳无踪迹，而正是这无影无踪的电子之魂在波动。从这个观点来看，一个电子就是电子迷雾中的一个尖峰。但人们很快就意识到，这种说法不可能正确，因为即使是一个尖锐的波峰——比如巨大的潮汐波——它最终也将传播开来。如果电子波会弥散开来，那么我们就可能在某个地方找到电子电荷的一部分或发现它质量的一部分。但是我们从未发现这样的情况。当我们找到一个电子时，我们总是发现所有的质量和电荷都集中在一个微小的、像点一样的区域内。1927年，马科斯·波恩提出了另一个看法，而这个观点的提出正是迫使物理学进入一个激进的新领域的过程中起决定性作用的一步。波恩提出，这种波既不是无影无踪的电子，也不是以前我们在科学中碰到的任何其他东西。波恩提出，这种波，是一种概率波。

为了便于理解，我们来看一张水波的快照，图中既有强度较高的区域（在波峰和波谷附近）又有强度较低的区域（波峰和波谷之间的传播区域）。水波的强度越大，对附近的船或海岸线造成的影响力就越大。波恩想象的波也有高密度区和低密度区，但是他赋予这些波的意义却令人大跌眼镜：空间中某点的波的大小与电子位于该点的概率成正比。最容易找到电子的地方就是概率波最大的地方，不容易找到电子的地方就是概率波较小的地方。而没有电子的地方，概率波在该点的概率就为零。

图4.5给出一张概率波的“快照”，图中说明中强调了波恩的概率诠释。不像水波的照片，这种图像实际上是不可能用照相机拍出来的。

没有人直接看到过概率波，而且根据传统的量子力学的解释，也没有人能看到。相反，我们应用数学方程（薛定谔，尼尔斯·玻尔，沃纳·海森堡，保罗·狄拉克以及其他人发展出来的）来计算出概率波在某些情况下的形状。然后再把理论计算结果与用下面的实验方法得出的结果相对比，计算出给定实验条件下的某个电子的假设概率波后，重复同一个实验多次，每一次都记录下所测得的电子的位置。不同于牛顿的预言，同样的实验同样的初始条件并不一定会得到同样的结果。相反，我们的测量产生了各种各样的测量位置。有时候电子在这儿，有时候电子又在那儿。如果量子力学正确的话，我们将发现电子在给定位置的次数与我们所计算出来的该点的概率波的大小成正比。80多年的实验显示，量子力学的预言有着非常高的精确度。

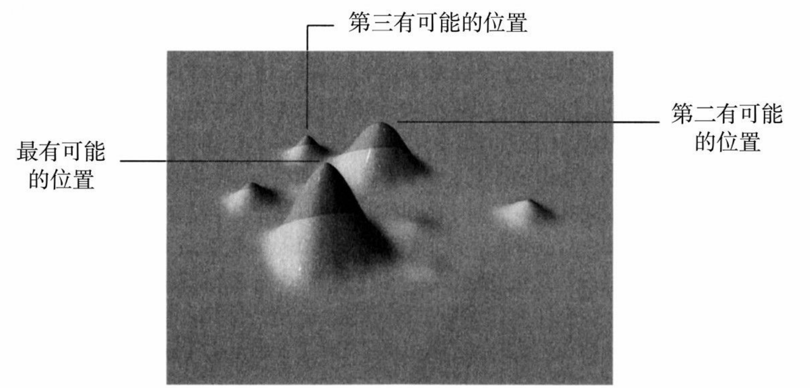



图4.5　粒子（比如说电子）的概率波告诉我们粒子出现在某个位置的可能性。

图4.5显示了电子概率波的一小部分：根据量子力学，每个概率波都可以扩展到整个空间，整个宇宙。[6]在许多情况下，一个粒子的概率波在某些小区域之外迅速衰减为零，表明该粒子出现在该区域内的可能性非常大。在这种情况下，概率波在图4.5外的部分（宇宙空间中的其他所有部分）看起来非常接近其在该图边缘部分的样子：相当平并且几乎接近零。然而，只要电子概率波的值在仙女座星系中的某处非零，不管该值有多小，我们总有一个微小但确实存在的——非

零——机会在那里找到电子。

因此，量子力学的成功之处在于迫使我们接受这样的观念：作为物质组成成分的电子，按我们通常的理解，占据着空间中很小的、类似于点的那么一点区域，实际上却也具有波的性质，并且可以弥漫到整个宇宙。而且根据量子力学，这种粒子波的融合适用于大自然中的所有成分，而不仅仅限于电子：质子同时具有粒子的特性和波的特性；中子也是如此，而且，早在20世纪初期就有实验表明，光——看起来是一种波，如图4.1——也可以从微粒的角度加以描述，我们在前面提到过小小的“光束”叫做光子。[7]打个比方来说，100瓦的灯泡所发出的一系列电磁光波，也可被描述成灯泡每秒发出一万亿亿个光子。在量子力学的世界里，每件事物都同时具有粒子属性和波的属性。

过去的80年中，量子力学的概率波在预言和解释实验结果上的普遍性和功用性已经得到了大家的信服。但量子力学的概率波究竟是什么，人们还没有达成一致的看法。我们应当说电子的概率波就是电子呢，还是只是说与电子有关呢，又或只是描述电子运动的数学工具呢，再或只是电子的某种具体表现形式呢？这一切都还在争论之中。但可以肯定的是，通过这些波，量子力学以人们无法预料的方式给物理定律注入了概率。气象学家用概率来预言降水的可能性，赌场用概率来预言你赢钱的可能性，但概率之所以会在上述例子中起作用，只是因为我们还没有足够的信息来做出确切的预言。根据牛顿所言，如果我们清楚地知道外部环境的每一点具体细节（每一个微小的组成成分的位置和速度），我们就可以确切地预言出（假定我们有足够的计算能力）明天下午4：07是否降水；如果我们知道有关骰子的所有物理细节（骰子的精确形状和成分，当骰子离开手时的速度和方向，桌子的成分和桌面情况等），我们将能准确预言出骰子落下时的具体形式。但从实际操作上来看，我们不可能收集所有的信息（即便我们能，我们也不可能有如此强大的电脑综合分析这么多数据，最后做出

预言），于是我们把眼光放低，在讨论有关天气方面或赌场方面的事情时，我们对我们所无法获得的数据做了一定合理的假设后，预言给定结果出现的概率。

量子力学中的概率完全是另一回事，在量子力学中，概率属于更为基本的特性。根据量子力学，不论我们的数据收集能力以及电脑运算力进步到何种地步，我们所能做的最好的事情就是预言这种或那种结果的概率。我们所能做的最好的就是预言电子，质子或中子，或其他物理学组成成分出现在这里或那里的概率。在微观领域里，概率至高无上。

再看图4.4的例子，量子力学关于单个电子逐个通过细缝形成图4.4所示的明暗相间的图样的解释，现在看来再清楚不过了。每个单独的电子都可以用其概率波加以描述。当一个电子射出时，它的概率波通过了两条缝隙。就像光波和水波那样，概率波通过这两条缝隙发散开来产生干涉。在接收屏的一些点上，两个概率波加强，使得强度变大。在另一些点上，两列波相互抵消，强度就低。在另外一些点上，概率波的波峰和波谷完全抵消，结果波的强度就为零。换句话说，电子落在探测屏上某些点处的概率较大，落在另外一些点处的概率较小，而有些点处电子则根本不会落在那里。经过一段时间，电子的落点就会按照概率图分布，因此我们就会在接收屏看到或明或暗乃至完全黑色的区域。详细的分析表明，这些亮区和暗区看起来正像图4.4所示的那样。

爱因斯坦和量子力学

由于固有的概率性质，量子力学完全不同于在其出现之前其他描述宇宙的理论，不论定性还是定量上都是如此。自其诞生之日起，物理学家们就在努力将这个奇怪的、出乎意料的理论体系融入普通的世界观中；这些努力到现在仍在进行中。问题在于如何协调日常生活的宏观经验与量子力学所展现的微观实在性。我们已经习惯于生活在这

个物理性质至少看起来稳固而可靠的世界上，尽管有时不得不屈服于各种各样的经济或政治突发事件的反复无常。你不用担心你现在呼吸进去的空气的原子会因其量子性质而突然解体，然后出现在我们看不到的月球那一面并重新组合起来，只留下你气喘吁吁地尽力呼吸着。你不用为这种事情而感到烦恼，因为根据量子力学，这种事情发生的概率即便不是零，也是相当小的。但是究竟是什么使得概率如此之小呢？

简单地说，有两点原因。第一，从原子尺度看，月球无限远。而且，正如我们在上文所提到过的，在许多情况下（虽然绝不是所有），量子方程显示，概率波一般只在空间中的很小区域内才有可测量的值，在一旦离开这个区域，概率波就会迅速衰减到几乎为零的程度（如图4.5所示）。因此和你出现在同一空间的单独一个电子——比如你刚刚呼出到空气中的一个电子——立刻出现在月球背面的可能性，即便不是零，也是非常小的。这个概率如此之小，以至于你娶到尼科尔·基德曼或嫁给安东尼奥·班德拉斯的概率都成了无比的大[13]。第二，在你房间的空气中，有许多电子、质子和中子。所有的粒子一起发生这种对一个粒子而言都是概率很小的事情的概率更是非常小，以至于都不值得我们思考。这就好比不仅要与令你心动的电影明星结婚，而且还要在相当长的一段时间内赢得每期的彩票，那么这个相当长的一段时间指的是多长呢？这个嘛，要长到能使现在的宇宙年龄看起来不过是宇宙眨了下眼。

这就使我们有点明白为什么在日常生活中我们不会直接遇到量子力学的概率方面的问题。然而，由于实验证实量子力学确实可以描述基础物理学，所以它对我们有关实在性的基本信念形成了威胁。比如爱因斯坦，就深受量子力学的概率性质所烦扰。他一再强调，物理学决定着什么已经发生，什么正在发生，以及我们周围的世界将会发生什么。物理学家不是赛马活动的下注者，物理学也不是用来计算概率的。但爱因斯坦不能否认量子力学在解释和预言微观世界的实验结果

方面非常成功，尽管这只是从统计学数据来看。因此爱因斯坦并未试图说明量子力学是错误的，因为考虑到它的成功之处那样做简直就是傻瓜才犯的错误，但他付出了很大的努力试图说明量子力学并不是决定宇宙如何运作的终极理论。虽然不能说明到底是怎么回事，但爱因斯坦试图向大家证实，一定存在着一种更为深刻又不那么诡异的有关宇宙的理论体系有待人们去发现。

前后许多年，爱因斯坦提出了一系列更加复杂的挑战，旨在找出量子力学结构上的缺陷。其中之一，在1927年索尔维研究所举办的第5届物理学会议期间提出，[8]基于这样一个事实：即使电子的概率波看起来如图4.5所示的那样，但不管何时，只要我们测量电子的位置，我们总是发现电子在或这或那的确切位置上。因此，爱因斯坦问道，这是否意味着概率波只是某种能够预言电子的确切位置的更为精妙的理论——只是尚未发现而已——的一个临时替身呢？毕竟，如果在X处发现了电子，那不就意味着，实际上电子在测量之前就在X处或非常接近X处？如果这样的话，爱因斯坦更进一步，量子力学所依赖的概率波——在本例中，表明电子距离X处有多远的概率的波——不正反映了理论在描述基础事实性方面的不充分性？

爱因斯坦的观点简单而又有说服力。认为粒子一直就在后来发现它的位置或邻近位置处岂不是非常自然的事情？如果是这样的话，物理学上更为深刻的理解需要我们能够给出那些信息，放弃粗糙的概率体系。但是丹麦物理学家尼尔斯·玻尔和量子力学的拥护者们不同意爱因斯坦的观点。他们认为，这样的推理源于传统的思维，即每个电子按照单独的、确定的路径往返。而图4.4强烈地挑战了这种思维，因为如果每个电子确实按照一个明确的路径——就像枪里发射出子弹这样的经典图像——那将很难解释观测到的干涉图样：是什么和什么发生干涉呢？正常情况下，从单独一支手枪一个接一个发射出来的子弹不可能与另一个子弹发生干涉，因此，如果电子像子弹一样运动的话，那图4.4的干涉图样又做何解释呢？

与之不同，根据玻尔和他力推的量子力学哥本哈根诠释，在人们测量电子的位置之前，问它在哪儿毫无意义，不会有一个明确的答案。概率波所记录的是，当适当地测量一个电子时，在或这或那的位置发现它的概率，而事实上这就已经是关于其位置所能讨论的全部了。没其他的了。只有当我们“看”电子的那一刻——我们测量其位置的时刻——并找到其具体位置时，它才具有通常意义上的确切的位置。但在我们测量之前（和之后），它所有的全部乃是概率波所描述的潜在位置，而概率波，就像其他任何波一样，具有干涉效应。这并不是说电子有一个位置，只是我们在测量之前不知道它的位置具体在哪儿。与你所想的恰好相反，在测量之前，电子根本就没有确切的位置。

这是一种从根本上来说非常奇怪的实在性。以这种观点看，当我们测量电子的位置时我们并不是在测量某种实体客观的、已经存在的性质。测量行为本身与创造出所要测量的实在性纠结在一起。将这种讨论从电子的尺度逐渐提升到日常生活，爱因斯坦讽刺道：“你真的认为，要是我们不看它，月球根本就没挂在天上吗？”量子力学的支持者们用森林中的落木的格言来回应他：如果没有人在看月亮——如果没有人“通过看见它来测量它的位置”——那么我们根本就不知道它是否在那儿，因而根本没有理由提出这样的问题。爱因斯坦认为这种说法令人非常不满意。这同他对于实在性的观点大相径庭；他坚定地认为，不管是否有人看见，月球总是挂在天空。但量子力学坚定的支持者们并没有被说服。

爱因斯坦的第二次挑战紧随其后，在1930年的索尔维会议中提出。他描述了一套假想中的装置（巧妙地结合了天平、时钟、相机似的快门），该装置看起来能证明电子一类的粒子必定有明确的性质——在测到它之前——而量子力学认为这根本是不可能的。细节倒不太重要，但解决方式十分具有讽刺性。当玻尔听说了爱因斯坦的挑战后，他十分震惊——乍看之下，完全无法在爱因斯坦的论证中找出一

点瑕疵。但是，几天之后，他卷土重来，完全驳倒了爱因斯坦的观点。令人惊奇的是，玻尔回应爱因斯坦的关键居然是广义相对论！玻尔意识到，爱因斯坦遗漏了他自己发现的引力扭曲时间——时钟运行的速度取决于它所感受到的引力场。当将这一复杂因素考虑进来后，爱因斯坦被迫承认他的结论与正统的量子理论完全一致。

即使他的异议被推翻，爱因斯坦仍然觉得量子力学令人非常不舒服。在后来的几年中，他一直盯着玻尔和他的同事们，不断提出一个又一个挑战。他最最富成效及具有深远影响的攻击集中在所谓的不确定原理上，这是沃纳·海森堡于1927年发现的量子力学的直接推论。

海森堡和不确定性

不确定原理就概率与量子宇宙之间的联系的紧密程度做出了准确的定量估计。为了便于理解，想想某些中国餐馆的菜单。菜肴编排在A、B两列中，举个例子来说，如果你点了A列菜单里的第一种菜肴，你就不能再点B列里的第一种菜肴；如果你点了A列里的第二种菜肴，你就不能再点B列里的第二种菜肴，如此等等。这样一来，餐馆已经建立起了饮食之间的对应关系，烹调间的互补（特别是那种设计出来防止你一次点很多昂贵菜肴的互补关系）。在菜单上你能点北京烤鸭或广东龙虾，但不能两个都点。

海森堡的不确定原理就是这样。粗略地讲，微观世界（粒子的位置、速度、能量、角动量，等等）的物理学特性能被分成A、B两列，就像海森堡所发现的，对A列第一种性质的了解会妨碍你对B列第一种性质的了解；对A列第二种性质的了解会妨碍你对B列第二种性质的了解；如此等等。而且，就像餐馆也会允许你同时点北京烤鸭和广东龙虾，但是却要在使其配比到总和不超过同样的总价的前提下，你对一列中的某种性质了解得越精确，你对第二列中的对应性质的了解就只能越不精确。从根本上说，海森堡发现的不确定原理说的就是从根本上不可能同时确定两列中的所有性质——即弄清楚微观世界里的所有

特性。

举个例子来说，你对一个粒子的位置知道得越精确，你对它的速度知道得就越少。同理，你对一个粒子的速度知道得越精确，你对该粒子的位置知道得就越少。量子力学有它自己的对偶性：你很精确地测定微观世界某种物理特性的同时，就失去精确测定另一些互补特性的可能性。

为了更好地理解，我们来看看海森堡自己的粗略描述，虽然其在某些特殊方面还不完善，但却能给我们有用的直觉图像。当我们测量物体的位置时，我们通常都会以某种方式与它发生相互作用。假如我们要在黑暗的房间里找到灯的开关，那么我们就得靠摸，摸到它就是找到了它的位置。如果一只蝙蝠在寻找一只田鼠，它就会发出声呐，并翻译出从目标身上反射回来的波。最为普通的定位方式就是看——物体上反射的光进入我们的眼睛。关键之处在于，这些相互作用不仅对我们有影响，也会影响我们要定位的物体。即使是光，从物体上反射回来时，也会对物体的位置有一个小小的影响。对于常见的物体——比如你手中的书或墙上的钟表——而言，反射光的轻微推力不会带来明显的影响。但当它撞击一个像电子一样的微小的粒子时就会产生巨大的影响：当光从电子表面反射回来时，它改变了电子的速度，就像你的速度会被一阵强烈的狂风影响一样。事实上，你越想精确的测定电子的位置，所使用到的光束的能量就得越强，对电子运动的影响也就会越剧烈。

这就意味着如果你高精度地测量电子的位置，就必然影响你的实验：精确定位这种行为本身就会影响电子的速度。因此你可以精确的定位电子，但是与此同时你却无法知道它的速度，因为那一刻它正在运动。相反，你能精确地测量出电子的运动快慢，但这样做的同时你会无法知道它的位置。大自然对确定这样的互补性质的精确度有一个内在的限制。虽然我们一直在讨论电子，但要知道不确定原理具有普适性：它适用于一切事物。

在日常生活中，我们常常会说些诸如一辆小汽车正以每小时90千米（速度）经过某车站站牌（位置），轻易地定下两种物理性质。事实上，量子力学认为这种说法没有精确的意义，因为你不可能同时测量确定的位置和确定的速度。对物理世界这样不精确的描述之所以对我们的生活没有影响，则是因为相比于日常生活的尺度，与不确定原理相关的尺度非常之小，以至于可以完全忽略不计。你看，海森堡原理指出存在不确定性，还指出了——完全确定地指出了——在任何情况下不确定性的最小量。假如我们把他的公式应用于你的小汽车经过一个车站时的速度，若对汽车位置的测量精确到厘米量级，那么速度的不确定性将还不到每小时万亿亿亿分之一英里。如果一名州警完全遵守量子物理原理，那他将会断言你在经过车站时的速度在每小时89.9999999999999999999999999999999999999和90.0000000000000000000000000000000000001英里之间。但是，如果我们用一个位置需要精确到十亿分之一米的电子来代替你庞大的汽车的话，那么它速度上的不确定度将是每小时100000英里。不确定性总是存在的，但是它只有在微观世界里才有意义。

对测量过程中不可避免的干扰的不确定性解释为物理学家们提供了一种有用的直观导引以及某些特定情况下强有力的认知框架。然而，它也容易造成误导。很容易造成这样一种印象，我们笨拙的实验员乱碰器材导致了不确定性的产生。这种想法是不对的。不确定性植根于量子力学的波动体系本身，不论我们是否做了笨拙的测量，它都一直存在。举个例子来说，我们来看一个特别简单的粒子概率波，有点类似于轻柔翻滚的海浪，如图4.6所示。既然波峰全都一致地向右移动，你可能会认为这列波描述的是以波峰的速度移动的粒子；实验证实了这种猜想。但是粒子在哪儿呢？因为这列波是均一地弥漫于整个空间，所以我们没有理由说电子在这儿或那儿。测量的话，我们会发现它无所不在。因此，当我们精确地知道粒子的移动速度时，它位置上的不确定性就会变得极大。如你所见，这样的结论与我们是否干扰了这个粒子无关，我们从未触碰过它。换句话说，它取决于波的基

本特征：波可以延展。

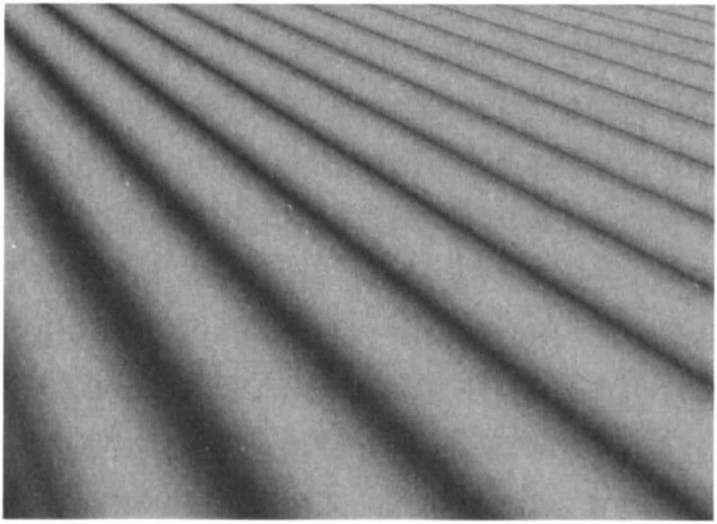


图4.6　具有均匀分布的波峰和波谷的概率波代表着粒子具有明确的速度，但由于波峰和波谷在空间中均匀分布，粒子的位置无法完全确定，它可能四处都在。

虽然细节上更麻烦一些，但类似的推理适用于其他所有波形，因而一般性的规律很清楚了。在量子力学中，不确定性必然存在。

爱因斯坦，不确定性和实在性问题

你可能正在被一个重要的问题烦扰着，即，不确定原理究竟是关于我们能够了解多少实在性呢，还是关于实在性本身呢？虽然理论上讲，量子不确定性告诉我们，我们永远无法同时搞清楚实在性的所有性质，比如位置和速度；但是，构成宇宙的基本成分，是不是像我们从日常事物——高飞的棒球，甬道上慢跑的人，向着太阳慢慢生长的向日葵——中得到的经典印象那样，真的有位置和速度呢？量子不确定性完全破坏了经典模型吗？它告诉我们，经典直觉赋予实在性的一系列属性，以构成世界的物质成分的位置和速度为代表的这一系列属性，是否会是一种误导？量子的不确定性是否告诉我们，在任何时

刻，粒子就是不能拥有确定的位置和速度？

对于玻尔而言，这个问题就像禅宗的以心传心一样。物理学家们只处理我们可以测量的事物。从物理的观点来看，这就是实在性。试图用物理学来分析“更深层”的实在性——那种远非我们可以通过测量来了解的实在性，就如同试图用物理学来研究孤掌鸣音一般。但在1935年，爱因斯坦和他的两位同事，鲍里斯·波多斯基和内森·罗森以一种有力而又巧妙的方式提出了这个问题，孤掌不再难鸣，他们3人抛出的问题对我们关于实在性的理解带来了连爱因斯坦也未曾预料到的冲击。

爱因斯坦波多斯基罗森论文是要证明，量子力学，尽管在做出预言及解释数据方面取得了巨大的成功，但并不能作为微观世界物理学的定论。他们的思路非常简单，基于刚刚提出的问题：他们想要证明，在任意给定的瞬间，每个粒子确实具有确定的位置和确定的速度；因而，他们想要得出不确定原理暴露了量子力学方法的根本局限性这样的结论。假如每个粒子确有位置和速度，而量子力学又不能处理这些实在的性质，那么量子力学就只不过是关于宇宙的部分描述。他们想要证明，量子力学只是一种不完善的物理实在性理论，或许，只是达到尚未发现的更深刻的理论框架的踏脚石而已。事实上，我们将会看到，他们的工作为某种更为奇妙的东西打好了基础，而这种东西就是：量子世界的非定域性。

爱因斯坦、波多斯基和罗森（EPR）的灵感部分是被海森堡对不确定原理的粗略解释激发的：当你测量物体的位置时你必然会扰动它，因而不可能同时测出它的速度。就像我们所看到的，虽然量子不确定性比“扰动”解释更具普适意义，爱因斯坦、波多斯基和罗森却从中发明了一套有说服力的巧妙办法避开任何不确定性的起源。他们提出，是否存在一种间接的测量方法可以同时测出粒子的位置和速度而不会影响粒子本身呢？举个例子，我们来看一个经典物理的类比，想象一下罗德和托德·弗兰德斯，他们决定在斯普林菲尔德新形成的核荒

漠中独自闲逛。他们返回到沙漠中心，背对背站立，约定朝着各自的方向以同样的速度一直向前走。一段时间，比如9个小时后，他们的父亲——内德——从斯普林菲尔德山返回的时候看到了罗德，他飞快地跑了过去，着急地问罗德，托德跑哪去了。这时候，托德已经走很远了，但是通过询问罗德相关情况，内德了解了有关托德的很多情况。如果罗德从开始的位置向东走了恰好45英里，则托德从开始的位置向西也恰好走了45英里。如果罗德以每小时5英里的速度向东走，托德也一定以每小时5英里的速度向西走。因此即使托德在90英里开外的地方，内德也能间接知道他的位置和速度。

爱因斯坦和他的同事把类似的策略应用于量子领域。在很多众所周知的物理过程中，两个粒子可以从共同的位置分开向相反的方向运动，就像罗德和托德一样。举个例子来说，如果一个单独的粒子分解成两个质量相等的粒子，它们分别朝“相反”的方向飞出去（就像一颗炸弹爆炸成朝着相反方向运动的两块），这是亚原子粒子领域的常见现象，两部分的速度相等，方向相反。而且，这两个组分粒子的位置也是紧密联系的，简单起见，我们总可以认为这两个组分粒子距离初始位置的距离相等。

有关罗德和托德的经典例子与上文所述的两个粒子的量子描述之间存在着重要的区别，即虽然我们可以肯定地说这两个粒子的速度之间存在着一种明确的关系——假如测量时发现其中一个以给定速度向左运动，那么另一个必然以相同的速度向右运动——但我们却不能预言粒子运动速度的具体数值。相反，我们所能做的就是运用量子物理的原理去预测粒子以某个速度运动的概率。类似的，尽管我们可以肯定地说这两个粒子的位置之间存在着一种明确的关系——假如在某一特定时刻我们测得一个粒子出现在某一位置，那么另一个必然会在距离出发点相等距离的位置上，只是方向相反而已——但我们却不能预言任何一个粒子的确切位置。相反，我们所能做的就是预言粒子在某个特定位置的概率。因此，量子物理并没有给出粒子速度和位置的明

确答案，它只是在某些特殊情况下，对粒子的速度和位置之间的关系做出明确的陈述。

爱因斯坦、波多斯基和罗森试图利用这些关系证明，每个粒子事实上在每个特定的时刻都有确定的位置和速度。这就是他们的想法：想象一下，你先测量了向右运动的粒子的速度，通过这种方式，就可以间接了解向左运动的粒子的速度。EPR认为，既然你没有对向左运动的粒子做任何事情，完全没有，那它就一定在那个位置上，你所做的只是确定它到底在哪儿，尽管只是间接的。然后，他们又巧妙地指出，你可以选择测量向右运动的粒子的速度。这样，你就可以完全不用打扰向左运动的粒子但却间接地测量出了它的速度。EPR再次指出，既然你没有对向左运动的粒子做任何事情，完全没有，它就一定以那个速度运动，你所做的只是确定这个速度是多少。将两者——你所做过的测量以及你可能已经完成的测量——结合起来，EPR得出结论：在任何给定的时刻，向左运动的粒子都有明确的位置和速度。

因为此处非常微妙且关键，所以我再说一遍。EPR的论证是，在你测量向右运动的粒子的过程中，你并没有对向左运动的粒子产生任何影响，因为它们是距离彼此很远的两个实体。向左运动的粒子完全不知道你对向右运动的粒子已经做的或可能做的测量。当你测量向右运动的粒子时，这两个粒子之间可能相距几米，几千米，或几光年远，因此，简而言之，向左运动的粒子不会在乎你究竟做了什么。因此，你通过研究向右运动的粒子的特性而了解的或者至少是理论上可以了解的有关向左运动的粒子的相应特性，一定是向左运动的粒子的确定的，一直存在着的特性，并且完全不会受你的测量影响。因此假如你测量了向右运动的粒子的位置，你将会知道向左运动的粒子的位置，同理，假如你测量了向右运动的粒子的速度，你将会知道向左运动的粒子的速度，这样，向左运动的粒子实际上有明确的位置和速度。当然，整个讨论将通过交换向左运动和向右运动的粒子角色而进行（而且，事实上，在做任何测量之前，我们根本没法说哪个粒子向

左运动，哪个粒子向右运动）；于是我们可以得出这样的结论：粒子具有确定的位置和速度。

就是这样，EPR认为量子力学只是实在性的不完备描述。粒子有确定的位置和速度，但量子力学不确定性原理却表明，实在性的这些特性远远超出了这个理论的掌控范围。假设你与大多数物理学家一样，相信自然界的完美理论应当能描述客观实在的所有属性，那么量子力学的在同时描述粒子的位置和速度上的失败，就意味着它不能描述某些属性，因此并不是一个完善的理论；量子力学不是自然界最终理论。这就是爱因斯坦、波多斯基和罗森严格讨论的问题。

量子回应

EPR认为，在任意给定的时刻，每个粒子都有确定的位置和速度，如果你顺着他们的过程进行的话，你将会发现，真的要确定下来这些性质将是不可能的。你可以按照上文所说的选择测量向右运动的粒子的速度，你这样做时将会扰乱它的位置；从另一方面来看，你测量它的位置的同时会扰乱它的速度。如果你不可能有向右运动的粒子的这些属性，你也不可能得到向左运动的粒子的这些属性。因此，不确定原理并没有矛盾：爱因斯坦和他的同事完全知道他们不能同时确定任意给定粒子的位置和速度。但是，问题的关键来了，即便没有同时确定其中任何一个粒子的位置和速度，EPR的论证表明，每个粒子都有确定的位置和速度。对于他们而言，这是一个实在性的问题。在他们看来，如果一个理论不能描述实在性的所有元素的话，它就不能声称自己为一个完善的理论。

量子力学的支持者们对EPR出乎意料的评价进行一番积极的回应后，继续专心于他们日常的实际研究，著名的物理学家沃尔夫冈·泡利总结道：“人们不愿意在那些你都不知道到底存不存在的事情上浪费脑力，就像没人愿意理那个古老的问题：针尖上究竟可以站多少个天使。”[9]一般来说物理学，尤其是量子力学，只能处理宇宙的可观测性

质。其他的东西显然不在物理学范畴内。如果你不能同时测量粒子的位置和速度，就没有必要讨论粒子是否同时有位置和速度。

EPR则不同意。他们坚持认为，实在性不只是探测屏上的读数；它的内涵远多于某一时刻所有观测现象的总和。没有人，绝对没有任何人，任何工具，任何仪器，任何物体一直在“看”月亮，但人们都相信，月亮挂在天空。他们仍然相信这只是实在性的一部分。

从某种意义上讲，这些讨论与牛顿和莱布尼茨之间对于空间实在性的争论遥相呼应。如果我们实际上没有办法摸到它，或看到它，或以其他的什么方式测量到它，我们应该认为这样的东西存在吗？在第2章中，我曾讲过牛顿的桶如何改变了有关空间的论战的性质，这个桶使人们突然意识到，空间的影响可以被直接观测到：桶中旋转的水会呈凹面。1964年，通过出人意料的突然一击——有评论者将其称为“最深刻的科学发现”[10]——爱尔兰物理学家约翰·贝尔也为量子实在性的论战做了同样的事情。

在后面的4节中，我们将讨论贝尔的发现，明智地避开了一点技术问题无疑是有好处的。尽管讨论所用的推理都不如计算掷骰子中的概率复杂，但却有一些我们必须讲述并将其前后串联起来的步骤。每个人对细节的要求不同，你可能读着读着就感到十分厌倦，那么没关系，请直接翻到本章最后，在那里可以找到有关贝尔的结论的讨论概要。

贝尔与自旋

约翰·贝尔把爱因斯坦—波多斯基—罗森论文的核心思想从哲学思考转化成了可以用具体的实验测量回答的问题。令人惊讶的是，达到这一点，他所需要的只是想出这样一种情形，在这里不只有两个由于量子不确定性原理而无法同时测定的性质——比如说位置和速度。贝尔证明，在不确定原理之下，如果同时有3个或更多的物理量——当

你测量其中的一个量时，将会影响其他物理量的测量，因此不可能同时测定这3种或更多种性质——则存在着一个能提出实在性问题的实验。这类例子中最简单的一个与所谓的自旋有关。

20世纪20年代开始，物理学家们已经知道了粒子的自旋——类似于橄榄球飞向目标的过程中会绕着自身旋转，粒子自旋就是这样的旋转运动。不过，量子粒子的自旋在一些重要方面不同于经典物理图像，对于我们而言主要是以下两点。首先，粒子——比如电子和质子——只能绕着特殊的轴以不变的速率顺时针或逆时针旋转；粒子的旋转轴能改变方向但旋转的速度不能减慢或加快。第二，量子不确定性也适用于自旋，正如你不能同时确定粒子的位置和速度，你也不能同时确定绕不同轴的自旋。举例来说，假如一个橄榄球绕着东北指向的轴旋转，那么它的自旋可以分解成绕向北和向东的轴的旋转——通过适当的测量办法，你可以确定下来在每个轴的方向上自旋的分速度。但是，如果你测量绕任意轴旋转的电子，你将无法得出自旋的分量。永远都不会。似乎测量本身会迫使电子将其在各个方向上的自旋运动集合起来，并且绕着你选定的轴顺时针或逆时针旋转。而且，由于你的测量影响了电子的自旋，你将无法确定它是怎样绕着水平轴，或是前后方向的轴，或是你测量之前的任何其他轴旋转的。量子力学自旋的这些性质很难完整的刻画出来，而这种困难无疑凸显了经典物理图像在揭示量子世界真实性质方面的局限性。不管怎样，量子理论的数学结构以及几十年来的许多实验，都使我们确信，量子自旋的这些特性是毋庸置疑的。

在这里介绍自旋的原因并不是为了深究粒子物理的复杂性。马上就会看到，粒子自旋的例子为抽取有关实在性问题奇妙而又出人意料的答案提供了很好的实验场。也就是说，虽然由于量子力学的不确定性，我们无法同时知晓绕多个轴自旋的情况，但是，一个粒子可以同时有绕每个轴旋转的确定分量吗？又或者不确定性原理是否告诉了我们其他一些东西呢？它是否要告诉我们，不同于实在性的经典概念，

一个粒子就是不会并且不能同时拥有这些性质呢？它是否要告诉我们，一个粒子处于被囚禁在量子监狱里的状态，没有绕着任何轴的确定自旋，除非有人或有某种东西测量它，引起它的注意，从而使其获得——取决于量子理论的概率——以或这或那的特定自旋值（顺时针或逆时针）绕着选定的轴旋转？通过研究这个问题，这个本质上与我们前面所探讨的粒子的位置和速度一样的问题，我们可以利用自旋来探索量子实在性的本质（寻找大大超越自旋这一特殊例子的准确答案）。我们现在来具体看一下。

就像物理学家戴维·玻姆明确证明的那样，[11]爱因斯坦、波多斯基和罗森的推理很容易推广到粒子绕任意选定的轴能否有确定的自旋这一问题上。下面具体看看。假设有两台探测器，它们能够测量入射电子的自旋，一台被安置在实验室的左边，另一台则在实验室的右边。假设有两个电子背对背从两个探测器中间的光源射出，而其自旋——而不是先前例子中的位置和速度——彼此关联。如何做到这些细节并不重要；重要的是能够这么做，事实上，很容易做到这点。我们可以如此设定其关联性，如果左右两个探测器要测量的是绕着指向相同方向的轴的自旋，它们将得到相同的结果：如果探测器要分别测量它们各自这边的电子绕垂直轴的自旋，那结果将是左边的探测器发现电子是顺时针旋转的，右边的探测器也将发现同样的结果；如果探测器测量的是绕垂直方向顺时针旋转60度的轴的自旋，左边的探测器测量到逆时针方向的自旋，而右边的探测器也将得到一样的结果；如此等等。又一次，在量子力学中，我们所能做的最好的事情就是预言探测器发现粒子顺时针或逆时针自旋的概率，但我们可以100%的肯定，不论一个探测器发现什么结果，另一个探测器也将会发现同样的结果。[14]

玻姆对EPR论证的改进之处在于，无论从哪个角度看，现在的讨论都与讨论位置和速度的原始版本一样。粒子自旋的关联性允许我们通过测量绕某轴向右运动的粒子来间接测量绕该轴向左运动的粒子的

自旋。由于测量是在实验室的右边完成的，所以无论如何都不可能影响向左运动的粒子。因此后者自始至终都有刚刚确定下来的自旋值；我们所做的只是测量它，尽管只是间接测量。而且，既然我们选择的是绕任意轴进行这一测量，那么相同的结论也应该适用于任意轴：即使我们一次只能确定绕一个轴的自旋，向左运动的电子绕每一个轴都有确定的自旋。当然，左边和右边电子的角色可以颠倒，从而使得每个粒子都有绕任意轴的确定大小的自旋。[12]

眼下看来，以自旋为例同以位置或速度为例没有明显的区别，你可能会和泡利的想法一样，认为没有必要思考这类问题。如果实际上你不能测量绕不同轴的自旋，那么去思考粒子是否绕每一个轴都有确定的自旋——顺时针或逆时针——有什么意义呢？量子力学，以及更广义上的物理，仅仅担负着解释可观测世界之性质的责任。而玻姆、爱因斯坦、波多斯基和罗森中，没有一个人认为测量是可行的。相反，他们所认为的是，粒子具有不确定原理所禁戒的性质，即使我们永远不能明确地知道这些性质的具体数值，粒子还是具有这些性质。这些性质是所谓的隐性性质，或者按更为普遍的说法则是隐变量。

这就是约翰·贝尔改变了一切的地方。贝尔发现，实际上你还是不能确定粒子绕多个轴的自旋，可如果绕所有轴真的都有确定的自旋的话，那么就存在着可检验的、可观测的自旋效应。

实在性检验

为了理解贝尔思想的要旨，让我们再来看看穆德和史考莉，想象一下，他们每人又收到了另外一个盒子，其中也有钛盒，但是有个重要的新特点。每个小钛盒不是仅有一扇门，而是有三扇门：一扇在顶面，一扇在侧面，另一扇在前面。[13]附带的信上说当盒子的任意一扇门打开时，每个盒子内的小球就会随机选择是发红光还是蓝光，如果某个盒子的不同的门（顶面，侧面，前面）打开，小球随机选择的颜色也有可能不同，但是一旦一扇门打开，小球发出了某种颜色的光，

那就再也没有办法确定另一扇门打开时将会发出什么颜色的光了。（在物理学的应用中，这些特点正符合量子力学的不确定性：一旦你测量了其中一个物理量，你就无法测量其他物理量了。）最后，这封信告诉他们：这两套钛盒之间还存在着一种神秘的联系：虽然这些盒子的门被打开时，盒子内的小球随机选择发出哪种光，但如果穆德和史考莉碰巧都打开了相同号码的盒子的同一扇门时，他们将会看到发出相同颜色的光。如果穆德打开了1号盒子的顶面的门看到发蓝光，信上预言史考莉打开她的1号盒子顶面的门时也会发现放出蓝光；如果穆德打开2号盒子的侧门时发出红光，那么史考莉打开她的2号盒子的侧门时也会看到红光，依此类推。确实，当穆德和史考莉打开前几个盒子时——通过电话商量每次打开哪个盒子上的哪扇门——他们证实了信上的预言。

虽然穆德和史考莉面对的是比先前更为复杂的情况，但乍看之下还是会发现，史考莉先前的推理还是适用。

“穆德，”史考莉说，“这和前些天的盒子是一个道理。只不过又来一遍，没有什么神秘的。每个盒子里的小球只是被预先设定好了程序。你不觉得吗？”

“但是现在有三扇门啊，”穆德用警告的口吻说，“这样的话，盒子里的小球是无法知道我们将要选择打开哪一扇门的，对吗？”

“它不需要知道我们打开的是哪扇门，”史考莉解释道，“那也是程序的一部分。我们来看个例子。拿着下一个还没打开的37号盒子，我也一样。现在，想象一下，为了便于讨论，假设我的37号盒子内的小球按这样的程序发光，比方说，如果顶门打开就会发红光，侧门打开就会发蓝光，前门打开就会发红光。我们把这个程序称作红、蓝、红。显然，如果送给我们盒子的人把相同的程序写进了你的37号盒子，那么，如果我们打开相同的门的话，我们就会看到发相同颜色的光。这就解释了‘神秘的联系性’。如果我们俩相同号码的盒子都是按

相同的指令运作，那么一旦我们打开相同的门的话，就会看到相同颜色的光。这实际上没有什么神秘的！”

但是穆德并不相信这些小球是按程序运行。他相信信上所说的。他相信当他们的盒子的门被打开时，盒子里的小球会随机选择红光或蓝光，因此他毫不怀疑地相信，他和史考莉的盒子间确实存在一些神秘的远程联系。

谁是对的呢？因为随机选择颜色过程中或之前没有办法检查盒子里的小球（记着，任何干扰都会造成盒子里的小球立即随机的选择发出红光或蓝光，从而无法弄清楚它是如何工作的），看起来似乎没有办法确证史考莉和穆德中到底谁才是对的。

但是，神奇的是，片刻思考之后，穆德意识到有个实验可以彻底解决这个问题。穆德的推理非常直接，但是需要有点我们在讲其他很多东西时很少使用的明晰的数学论证。毫无疑问，这里的细节还是值得学习的——并没有多少——如果觉得稍有麻烦，也不用担心；我们很快就会总结关键性结论。

穆德意识到，迄今为止，他和史考莉只知道他们打开给定号码的盒子的相同的门时会发生什么。电话一接通，他就兴奋地告诉史考莉，如果他们不打开相同的门，而是随机且各自独立的打开每个盒子上的任意一扇小门，他们可能会了解得更多一些。

“穆德，拜托。请让我享受自己的假期吧。那样做我们又能知道什么呢？”

“喂，史考莉，我们能确定你的解释是正确的还是错误的。”

“好，我听着呢。”

“很简单，”穆德继续道，“如果你是对的，按我的认识我们就应该

发现：如果对于给定盒子，我们单独随机的选择打开某一扇门，并记录下它们发光的颜色，那么所有的盒子都打开后，我们就应该发现，我们俩看到相同颜色的概率在50%以上。如果不是这样的话，我们看到相同颜色的光的概率不足50%的话，那你的说法就是错误的。”

“是吗，怎么回事？”史考莉有点儿感兴趣了。

“好，”穆德继续道，“举个例子来说。假设你是对的，每个小球按照程序运作。具体一点，假设在某个盒子里小球的程序是蓝、蓝、红。因为我们都是从三扇门中选择，所以我们会有9种打开盒子门的组合方式。比如，我选择打开盒子顶上的门，而你选择打开侧门；或者我打开前门，你选择打开顶面的门；等等。”

“是的，当然，”史考莉高兴地说，“如果我们把顶门称作1，侧门称作2，前门称作3，那么9种可能的组合就是（1，1），（1，2），（1，3），（2，1），（2，2），（2，3），（3，1），（3，2），（3，3）。”

“嗯，是这样的，”穆德继续道，“现在这是关键：在这9种可能性中，注意其中的5种组合——（1，1），（2，2），（3，3），（1，2），（2，1）——将会使我们看到盒子里的小球发出相同颜色的光。对于前三种门的组合来说，是我们碰巧选择了相同的门，就我们所知，那会导致看到相同颜色的光。其他两种门的组合，（1，2），（2，1）也会发出同样颜色的光，因为，如果门1或门2被打开的话，程序就会指定这些小球发相同颜色的光——蓝色。因为5大于9的一半，这就意味着在多半——大于50%——的概率下，我们开门时会看到盒子里的小球发出同样颜色的光。”

“但是，等等，”史考莉抗议道，“这只是一个特殊程序的例子：蓝、蓝、红。在我的解释中，不同号码的盒子可能有不同的程序。”

“事实上，那没有关系。这个结论适合于所有可能的程序。你看，我对蓝、蓝、红程序的推理只依赖于这样一个事实，即程序中有两种颜色是相同的，相同的结论也适合于其他程序：红、红、蓝，或者是红、蓝、红，等等。任何程序中都至少有两种相同颜色；唯一不同于其他程序的是三种颜色都相同的程序——红、红、红和蓝、蓝、蓝。对于按这两种程序运作的盒子来说，不管我们碰巧打开的是哪扇门，我们都会看到同样颜色的光，这样我们俩看到相同颜色光的概率就会增加。因此，如果你的解释是对的，盒子按照程序运行——即使所有编号的盒子的程序都不相同——我们看到相同颜色的光的概率就应该大于50%。”

这就是他的论证。困难的部分已经结束了。最后，的确有一种验证史考莉的说法是否正确，每个小球是否按照程序——程序决定了打开哪一扇门时发出哪一种颜色的光——运行的办法。如果对每个盒子，她和穆德都各自独立又随机的打开三扇门中的一扇，然后对比他们各自看到的颜色——按照号码一个盒子接一个盒子的对比下去——他们一定会发现看到相同颜色的光的概率在50%以上。

正如在下一部分将要讨论的那样，用物理学语言说，穆德的发现正是约翰·贝尔的发现。

用角度来数天使

我们可以直接将这个结果翻译成物理语言。想象一下，我们有两个探测器，一个在实验室的左边，另一个在实验室的右边，这两个探测器可以测量诸如电子之类的粒子的自旋，就像上上小节所讨论过的实验一样。你需要为探测器选择测量哪根轴（垂直的、水平的、前后的，或者它们之间的无数其他轴）的自旋；为简化起见，可以这么想，出于预算的考虑，我们的探测器只能选择3种轴。不论实验如何运行，你都会发现电子只能绕着你选择的轴的顺时针旋转或逆时针旋转。

根据爱因斯坦、波多斯基和罗森的想法，每个入射电子带给探测器的信息构成一个程序：即使自旋这样的信息隐藏起来了，即使你不能测量到自旋，EPR还是认为电子会以确定的自旋——顺时针或逆时针——绕每个轴。因此，当电子进入探测器时，电子无疑会确定下来其绕着你所随机选定的轴的自旋究竟是顺时针还是逆时针。举例说明，绕着3个轴都顺时针旋转的电子所构成的程序就是：顺时针，顺时针，顺时针；绕前两个轴顺时针旋转，第3个轴逆时针旋转的电子带来的程序为顺时针，顺时针，逆时针，如此等等。爱因斯坦、波多斯基和罗森解释向左运动的电子和向右运动的电子之间的关联性的说法非常简单，他们认为，这些电子具有相同的自旋，因而带着相同的程序进入了探测器。因此，如果左边和右边的探测器选择了相同的轴，那么两台自旋探测器上得到的结果将会一模一样。

要知道这些自旋探测器完全等价于穆德和史考莉遇到的问题，只需做简单的替换即可：我们选择的不是钛盒上的门，而是选择轴；看到的不是发红光或蓝光，而是记录顺时针还是逆时针自旋。因此，正如打开一对相同号码的钛盒上的同一扇门时会看到相同颜色的光，两台选择了相同轴的探测器也将会得到同样的自旋结果。正如打开钛盒上的某一扇门会使我们无法知道若打开的是另一扇门的话会发出什么颜色的光一样，根据量子不确定性，测量电子绕某轴的自旋会使我们无法知道若选择另一根轴的话电子的自旋会如何。

上述讨论意味着穆德在判断谁才正确时所用的分析方法也同样适用于这里。如果EPR是正确的，如果每个电子绕3个轴都有确定的自旋——如果每个电子都自带“程序”，可以明确地定出3种可能的自旋测量的结果应该为何——那么我们就能做出下列的预言。仔细审查从多次实验——每次实验中，探测器都各自独立的随机选择轴——中收集来的各种数据，我们应该会发现，两个电子的自旋一致，都是顺时针或都是逆时针的概率大于50%。如果两个电子自旋不一致的概率并没有大于50%，那么爱因斯坦、波多斯基和罗森的观点就是错误的。

这就是贝尔的发现。他的实验表明，即使你不能实际测出电子绕多个轴的自旋——即使你没法明确地“读出”电子进入探测器时的程序——那也并不意味着试图弄清楚电子在每个轴上是否都有确定的自旋这样的问题，等同于搞清楚针尖上可以站多少个天使。远远不是这样。贝尔发现，粒子是否可以在每根轴上都有确定的自旋这个问题有一个真正的、可通过实验验证的结果。利用不同角度的3根轴，贝尔找到了一种数清泡利的天使的办法。

不期而至

即使你略过了许多细节也没有关系，我们来总结一下。根据海森堡的不确定性原理，量子力学宣称这个世界的一些性质——比如粒子的位置和速度，或粒子绕不同轴的自旋——不能同时有确定的值。根据量子理论，粒子不可能具有确定的位置和确定的速度；粒子不可能沿多根轴都有确定的自旋（顺时针或逆时针）；粒子不能同时拥有在不确定性下处于对立面两端的确定属性。相反，粒子处于不稳定的量子状态中，总是处于各种不同状态按概率的叠加中；只有在被测量的时候才会从众多状态中挑出一种确定的结果。显然，这与经典物理学所勾画的客观实在性大相径庭。

出于对量子力学的怀疑，爱因斯坦及其合作者波多斯基以及罗森，试图将量子力学的这个方面当成反对量子理论本身的武器。EPR认为，即使量子力学不允许同时测定所有这些性质，粒子却可以有确定的位置和速度，粒子在所有轴的方向上也可以有确定的自旋；对于所有的性质，粒子都可以具有量子不确定性所禁戒的确定值。因此，EPR认为，量子力学并不能处理物理上的客观实在的所有元素——它不能处理一个粒子的位置和速度；它不能处理一个粒子在多个轴上的自旋——因此量子力学不是一种完善的理论。

相当长的一段时间内，关于EPR是否正确的问题看起来更像是一个宇宙哲学问题而不是物理学问题。但正如泡利所言，如果你在实践

中不能测量量子不确定性所禁戒的性质，如果这些性质隐藏在物理实在背后，那么这同它们根本就不存在又有什么不同呢？但是，令人惊奇的是，约翰·贝尔发现了爱因斯坦、玻尔以及20世纪其他理论物理学巨匠没有发现的事情；贝尔发现，尽管我们无法通过测量获知某些事物是否真的存在，但如果它们存在的话，那就真的会带来一些不同之处——而这些不同之处可以用实验来验证。贝尔认为，如果EPR是正确的，那两个间隔很远的探测器在测量某些粒子性质（在我们所用的方法中，就是要测量绕各种随机选择的轴的自旋）时所得到的结果，彼此相符的概率将在50%以上。

贝尔在1964年就有过这种设想，但当时的技术还不足以完成这样的实验。到了20世纪70年代早期，技术上的障碍消除了。先后有很多人都做过这个实验，最早的有伯克利的斯图尔特·弗里德曼和约翰·克罗萨，紧接着有德州农工大学的爱德华·福莱和兰德尔·汤普森，并在20世纪80年代早期通过法国的艾伦·埃斯拜科特和他的同事们的工作得以完善，后来又涌现出了很多更为精炼及令人印象深刻的版本。我们以埃斯拜科特的实验为例，两个探测器间隔13米之远，装有高能量态的钙原子的容器放在它们之间。根据已经非常成熟的物理学知识，当钙原子回到正常状态，即较低能量状态时，将会射出两个背对背的光子，其自旋具有完美的关联性，就像我们先前讨论过的电子之间的自旋关联一样。确实，在埃斯拜科特的实验中，只要这两个探测器的设置一样，两个发射出的光子的自旋就会表现出完美的关联。如果我们还用发光来说明埃斯拜科特的实验，发红光对应顺时针的自旋，发蓝光对应逆时针的自旋，那么入射光子将会使探测器发出相同颜色的光。

但是，这正是关键之处，当埃斯拜科特查看多次运行之后所得出的大量实验数据时——左右探测器的设置并不总是一样的，而是随机且各自独立设置的——他发现两个探测器彼此符合的概率并没大于50%。

这个结果令人异常震惊，这就是那种令人惊讶到不能呼吸的结果。但为防止没说清楚，让我再来进一步解释一下。埃斯拜科特的实验结果表明，爱因斯坦、波多斯基和罗森等人的想法被实验——不是被理论，不是被思考，而是被自然——证明为错误的。而这就意味着，虽然EPR通过论证得出这样的结论：一个粒子可以具有哪些为不确定性原理所禁戒而不能有确定值的物理性质——像绕多个不同轴的自旋——的确定值，但是，在他们的论证过程中存在错误。

但是他们哪儿错了呢？这个嘛，别忘了爱因斯坦、波多斯基和罗森的论证中有一个核心假设：如果你想在某个给定时刻测量某一物体的性质，那么你可以测量空间上距离该物体很远的另一物体的性质，而前一物体总是具有这种性质。他们的这个假设非常简单且合乎道理。你的测量在此地进行，而前一物体在彼处。这两个物体在空间上相距很远，因此你的测量不可能对前一物体有任何影响。更准确地说，既然没有物体比光的速度更快，那么，即便你对某一物体的测量不知何故影响了另一个物体——比如说使另一个物体也做同样的绕选定轴的自旋运动——在这种影响发生之前也会有一个时间上的延迟，这个延迟时间至少得长到光可以穿越这两个物体之间的间隔。但无论是在我们的抽象论证还是在实际验证中，这两个粒子都是同时被探测器探测到。因此，我们通过测量第二个粒子所得到的关于第一个粒子的性质，一定就是第一个粒子实际具有的性质，这一点与我们如何进行测量毫无关系。简而言之，爱因斯坦、波多斯基和罗森的观点的核心是彼地的物体不会在乎你对此地物体所做的事情。

但正如我们刚刚看到的那样，这样的论证会带来如下的预言：两台探测器在多数情况下都会发现同样的结果，可这样的预言又被实验结果否定了。这样一来，我们就被迫得出如下结论，即，不管爱因斯坦、波多斯基和罗森所做的假设看起来多么合理，都不可能是量子宇宙的运行原理。因此，在经过这种间接却经得起推敲的思考之后，实验使我们得出这样的结论：彼地的物体确实在意你对此地物体所做的

事情。

根据量子力学，粒子在被测量时会随机获得这种或那种性质，而这种随机性，我们现在知道，可以超越空间联系起来。恰当制备好的成对粒子——所谓的纠缠粒子对——不会各自独立地获得他们的测量性质。他们就像一对魔法骰子，一个在大西洋，另一个在拉斯维加斯，每一个骰子上的数字都随机出现，但这两个骰子却不知为何总能保持一致。纠缠粒子对就是这样，只不过它们依靠的不是魔法。纠缠粒子对，即使空间上相隔很远，也不会自主运行。

爱因斯坦、波多斯基和罗森努力证明量子力学并不是有关宇宙的完备理论。半个世纪过去了，由他们的工作所引发的理论思考和实验结果却要求我们转回到他们全部论证的开端，我们最终发现他们的论证中最基本的、合乎直觉的、最有道理的那部分竟是错的：宇宙并不具有定域性。即使没有物体在这两个地点之间传播——即使没有足够的时间让物体在两个地点之间往来，你在一个地方做的事情还是会和另一个地方发生的事情有所关联。爱因斯坦、波多斯基和罗森那直观看来令人放心的想法——之所以有长程关联存在，只不过是因为粒子有确定的、先前就已存在的关联性质——被实验数据排除掉了。这就是为什么这一切如此令人吃惊。[14]

1997年，日内瓦大学的尼古拉斯·吉辛和他的研究组做了另一个版本的埃斯拜科特实验，两个探测器被置于间隔11千米远的两个地方，结果并未改变。相对于光子波长的微小尺度，11千米简直大到难以想象。对于光子而言，那简直是11000000千米，或是1100000000光年。我们有理由相信，不管探测器相距多远，光子之间的关联性总会存在。

这听起来非常奇怪。但现如今，所谓的量子纠缠早有了确凿的证据。如果两个光子处于纠缠状态，那么成功的测量其中一个光子绕轴的自旋将会“强迫”另一个远方的光子以同样的大小绕相同的轴自旋；

测量一个光子会“迫使”另一个远方的光子挣脱概率的迷雾获得确定的自旋值——该值会与远方的光子自旋精确匹配。这真的让人困惑不已。[15]

纠缠与狭义相对论：标准观点

我之所以会在文章中使用“强迫”和“迫使”这两个词，则一方面是因为这两个词传递出了我们那经典物理式的直觉所需要的情绪，而另一方面更是因为其在这里的精确意思对我们是否能很好地接受更为猛烈的观念转变至关重要。这些词在日常生活中的定义使我们在脑海中勾勒出一幅有关意志的因果关系图：我们在此地做一些事情会造成或迫使彼地发生一件特殊的事情。如果这就是有关两个光子之间如何纠缠的正确描述，狭义相对论就危险了。实验表明，从实验者的角度看，一个光子的自旋被测定的一刹那，另一个光子就会立即获得相同的自旋性质。如果有什么东西从左边的光子飞到右边的光子，告诉右边的光子左边光子的自旋已经被测定，那么它必须得在两个物体之间瞬间移动，而这就会与狭义相对论的速度极限相矛盾。

物理学家们大都认为，此类与狭义相对论有明显冲突的说法是错误的。直观上的原因是，即使两个光子间隔很远，它们的共同起源也会在它们之间建立一种基本联系。虽然它们彼此相背而行，在空间中分离，但它们的历史使它们纠缠在一起；即使相隔很远，它们仍然是整个物理系统中的一部分。这样的话，对一个光子的测量确实不会强迫或迫使另一个遥远的光子具有相同的特性。更确切地说，由于这两个光子关系密切，我们因此可以合理地认为它们——即使空间上相隔——是整个物理实体的一部分。于是我们就可以说，对一个独立整体的测量——该整体包括两个光子——会影响整体；也就是说，它会同时影响两个光子。

虽然这种说法可以使两个光子之间的联系更便于理解，但表述上却不清不楚——把两个空间中相隔很远的物体看做一个整体究竟是什

么意思？更准确的说法应该像下面这样。狭义相对论认为没有物体能比光传播得更快，“物体”指熟悉的物质和能量。但现在的情况非常微妙，因为似乎并没有物质和能量在两个光子之间飞行，因此也就没有必要测量速度了。不过，有一种方式可以让我们知道这是否与狭义相对论相冲突。物质和能量的一个特点是，从一个地方传播到另一个地方时会传递信息。光子从广播站传播到你的收音机是为了传递信息。电子从因特网电缆中传播到你的电脑是为了传递信息。因此，在任何情况下如果有某物体——即使是未命名的物体——声称比光传播得还快，最好的检验方法就是看它是否传递或至少能传递信息。如果答案是“不”的话，标准的论证就可以继续，没有物体的速度可以超越光速，狭义相对论仍然未被挑战。实际上，物理学家们经常用这个方法来检验某些微妙的过程是否违背了狭义相对论原理（还没有哪个过程能通过这个验证）。这里我们也来使用一下这个方法。

有没有这样一种方法，通过测量沿某轴向左和向右运动的电子的自旋，我们就可以在两个光子之间传递信息？答案是否定的。为什么呢？这个嘛，左边或右边探测器所得到的数据只不过是由顺时针或逆时针结果组成的随机序列，因为在任何一次实验运行中，粒子选择不同自旋的概率都是一样的。我们没有办法控制或预言某次测量的结果。因此，在这两个粒子随机选择的结果中，没有任何信息，没有隐藏的密码，什么都没有。唯一有趣的事就是这两列结果完全一样，不过，只有把它们放在一起，用传统的比光还慢（传真、电子邮件、电话，等等）的方法作比较时，我们才能看出这一点。因而，根据标准的观点我们可以得出这样的结论：虽然对一个光子自旋的测量貌似会立即影响另一个光子，但它们之间并没有信息传递，狭义相对论的速度极限仍然有效。物理学家们的确说了自旋具有关联性——因为探测器上所得到的两列结果是一样的——但这并不代表传统意义上的因果关系，因为在这两个相距很远的光子之间没有物体传播。

纠缠和狭义相对论：反方观点

真是这样吗？量子力学的非定域性和狭义相对论之间潜在的冲突完全解决了吗？或许吧。在上述思考的基础上，绝大多数物理学家认为狭义相对论和埃斯拜科特关于纠缠粒子对的实验可以和谐共存。简而言之，狭义相对论好不容易才过了这一关。许多物理学家觉得欢欣鼓舞，但有些人挑剔地认为还隐藏有更多的科学本质未被发掘出来。

在内心深处我总是持共存的观点，但并不否定这个问题非常棘手。到了最后，不管有人多么强调整体性，不管有人多么强调缺乏信息，两个相距很远的粒子，都是由量子力学的随机性主宰的，却总是能够保持“联系”，无论其中一个做什么，另一个都会立刻跟着做。所有的这些似乎都在暗示我们，有某种比光还快的物质操纵着它们。

那么我们应该持什么样的观点呢？这个问题没有严格的、被广为接受的答案。一些物理学家和哲学家们认为，要想有所进展，就得认识到迄今为止的讨论焦点有所偏差：他们正确地指出了狭义相对论的真正核心，并不在于光设定了一个速度极限，而是在于所有的观测者，不管是否处于运动状态，对于光速的认识完全一致。[16]概括地讲，这些研究者们强调，狭义相对论的核心在于，没有哪个观测者处于超越其他观测者的优势地位。因此，他们提出（许多人也同意），如果平等对待所有匀速观测者与纠缠粒子对的实验结果不相矛盾，那么人们对于狭义相对论的疑虑将得到解决。[17]但要达到这个目标并不是一项容易的工作。为了了解得更为具体，我们先来看看过时的量子力学课本怎样解释埃斯拜科特的实验。

根据标准的量子力学，当我们测量某个粒子，并发现它在这里时，会造成其概率波的改变：潜在的可能结果减为实际测量中看到的那一种，如图4.7所示。物理学家们认为测量会造成概率波塌缩，而且，初始概率波在某个位置处越大，概率波塌缩到该处的概率就越大——就是说，粒子出现在该点的概率就越大。在标准方法中，塌缩过程在整个宇宙中瞬间发生：一旦你在这里发现了粒子，在别处找到它的概率就立即减为零，这就是概率波瞬间塌缩的反应。

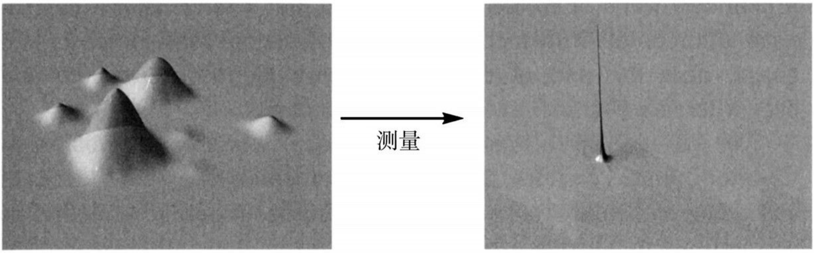



图4.7　当一个粒子在某个位置被观测到时，在其他位置发现它的概率将衰减为零，而在该位置观测到它的概率将变为100%。

在埃斯拜科特的实验中，如果测到了向左运动的光子的自旋，比如说发现它绕某根轴顺时针自旋，那么它在整个空间内的概率波就会塌缩，而逆时针旋转的概率则会瞬间归零。既然塌缩无处不在，它也就有可能出现在向右运动的光子所在的位置。而且，向右运动的光子概率波的逆时针旋转部分也会受到影响，同样塌缩为零。因此，不论向右运动的光子距离向左运动的光子有多远，它的概率波都会瞬间受到向左运动的光子概率波变化的影响，进而保证它和向左运动的光子一样绕某特定的轴作自旋。于是，在标准的量子力学中，造成比光速还快的影响的就是概率波的这种瞬间改变。

量子力学的数学使得定性讨论精确起来。而且，实际上，来自于概率波塌缩的长程影响改变了埃斯拜科特实验中左右两个探测器（它们的轴是随机独立选择的）出现相同结果的概率的预言。借助于数学计算才能得到精确的答案（如果感兴趣的话请参考注释部分[18]），数学计算的结果预言，探测器刚好有50%的一致率（而不是以前预言的大于50%的一致率——这个结果，我们已经看到，正是利用EPR的定域宇宙假设才得到的）。难以相信的精确，这正是埃斯拜科特在他的实验中所得出的结果，50%的一致率。标准量子力学竟与实验数据如此的匹配！

这一成就引人注目。不过，还有一个问题。70多年过去了，无人能理解概率波的塌缩是如何发生的，甚至根本无人知道概率波的塌缩

能否真正发生。这些年来，概率波塌缩假说把量子理论预言的概率同实验中得到的确定结果有力地联系起来了。但是，概率波塌缩假说本身就是个谜。首先，塌缩并不能由量子理论的数学推出；它是人为放进理论中的，而且也没有妥当的实验方法来验证。其次，我们在纽约的探测器中发现一个电子，结果造成了该电子在仙女座星系中的概率波瞬间归零，这怎么可能呢？当然，一旦你在纽约发现了某粒子，你就不可能再在仙女座星系找到它了。但是，究竟是什么不为人知的机制促使这样的奇迹成真呢？或者，更加形象地说，概率波在仙女座星系的部分，以及在其他任何地方的部分，究竟是怎样“知道”要同时衰减为零的呢？[19]

我们将在第7章中讨论这种量子力学的测量问题（我们将会看到，人们提出了其他一些避免概率波塌缩的观点），但在这儿，只要注意到下面的事情就足够了，正如我们在第3章中所讨论过的那样，从某个角度来看是同时发生的事情，从另一个运动着的视角来看就不是同时发生的了（想象一下傻猫和坏鼠在开动的火车上设置钟的例子）。因此，从某观测者的角度来看概率波在整个空间同时塌缩，但从另一个运动着的观测者的角度看则不是同时发生的。事实上，由于观测者的运动，一部分观测者告诉我们左边的光子先被测量到，而其他观测者则告诉我们右边的光子先被测量到，他们全都值得信赖。因此，即使概率波塌缩的假说是正确的，我们也没有一个客观的标准用以断定到底是哪一边——左边或右边的光子——的测量影响另一边。因而，概率波的塌缩似乎挑出了一个特别的观测点，即能使波函数的塌缩在整个空间中同时发生的点，能使左边和右边的测量同时进行的点。但选取一个特殊的视角就与狭义相对论核心的平等主义相矛盾。人们已经提出一些方案来规避这一问题，但是，到底哪一种，或哪些种观点才是正确的论战仍在继续。[20]

因而，虽然大多数人认为量子力学、纠缠粒子态和狭义相对论可以和谐共存，但也有一些物理学家和哲学家认为它们之间的确切关系

仍是一个悬而未决的问题。有可能，在我看来甚至是极有可能，多数派的观点更有可能以某种确定的方式获得最终的胜利。但历史告诉我们，奥妙而基础的问题有时会播下未来解决的种子。关于这点，只有让时间来验证了。

## 我们将用什么来解释这一切

贝尔的推理和埃斯拜科特的实验表明，爱因斯坦脑中的那种宇宙只存在于思想中，而不是现实中。在爱因斯坦预言的宇宙中，你在彼地所做的事情只会立即影响彼地的事物。物理，在他看来，是纯定域性的。但我们看到，从实验中得到的数据排除了这种想法，进而排除了这种宇宙。

在爱因斯坦预言的宇宙中，物体所有可能的物理性质都具有确定的值。物理性质并非空中楼阁，需要实验家们的测量才能变得实在。大多数物理学家可能会说爱因斯坦在这一点上也是错误的。在大多数人看来，粒子的性质只有在测量的驱使下才会实在起来——我们将会在第7章中进一步探究这一思想。当粒子未被观测或与环境没有相互作用时，粒子特性将会处于一种模糊、混乱的状态，该状态只能用这样或那样的可能成真的概率来描述。持该观点的最极端的那些人甚至声称，当没有人或没有物体“看”月球或与月球以某种方式相互作用时，它根本就不在那儿。

关于这个问题，现在仍有很多争议。爱因斯坦、波多斯基和罗森认为，关于测量竟能够发现空间上相隔的粒子具有相同性质的唯一合理解释是，粒子一直就具有那些确定的性质（并且，由于它们共同的过去，它们的性质彼此关联）。几十年过去了，贝尔的分析和埃斯拜科特的实验数据表明，这种建立在粒子总是有确定性质的基础上的直观上令人满意的说法，无法解释实验上观测到的非定域性关联。但无法解释非定域性的神秘性，并不意味着粒子总有确定的性质的说法本身是错误的。实验数据虽然排除了定域性宇宙，但并未排除粒子具有

这样的隐性质。

事实上，在20世纪50年代，玻姆创立了个人版本的量子力学，其中包含了非定域性和隐变量。按玻姆的理论，即便我们不能同时测量，粒子也总是有确定的位置和速度。玻姆的理论所做出的预言与传统量子力学的预言完全一致，但他的理论引进了一种更为明显的非定域性元素——作用于某一位置的粒子的力瞬时依赖于遥远的另一位置处的物理条件。从某种意义上来讲，玻姆版本的量子力学告诉我们的是，向着爱因斯坦恢复被量子革命所摒弃但直观上合理的某些经典物理性质——粒子有确定的性质——这一目标可以前进多远，但它也同时告诉我们，这样做的代价是接受更为夸张的非定域性。付出了这样沉重的代价，恐怕爱因斯坦很难找到一点安慰了。

我们从爱因斯坦、波多斯基、罗森、玻姆、贝尔和埃斯拜科特，以及在该研究方向上曾起过重要作用的其他许多人的工作中所学到的最令人惊讶的一课，就是需要摒弃定域性。由于它们的过去，现在遍布于宇宙不同区域的物体可能是量子力学纠缠整体的一部分。即便空间上相隔很远，这些物体仍然以随机而协调的方式演化。

我们曾经认为空间的基本性质在于分离区分物体。但我们现在看到，量子力学强烈地挑战了这个观点。两个物体可以在空间相隔很远，但却并不是完全独立存在。量子关联可以把它们统一为一个整体，使其中一个的性质取决于另一个的性质。空间并不能阻碍它们之间的相互联系。空间，即使是巨大的空间，也不能削弱它们之间量子力学导致的相互依赖性。

有些人把这解释为“每件事物总是与其他事物相关联”或“量子力学使我们生活在一个整体中”。毕竟，继续思考的话，所有物体在宇宙大爆炸时都来源于一个地方，我们相信，所有我们现在认为不同的位置都可追溯到一个起源。因此，像源于同一钙原子的两个光子一样，每个物体从起源上都源于一个物体，每个事物从量子力学上看都与其

他物体纠缠在一起。

虽然我偏爱这种观点，但这样富于感情色彩的说法不严密而且有些言过其实。源于钙原子的两个光子之间的量子关联确实存在，但却异常精巧。埃斯拜科特和其他人做他们的实验时，很关键的一点就是光子必须从发射源毫无阻碍地到达探测器。要是光子在到达探测器前与乱溅的粒子碰撞或与仪器的各部分相撞，则光子之间的量子关联将变得难以确定。那样一来，我们需要做的就不仅是找到两个光子性质之间的关联性，而是要找到光子和它可能碰撞的其他物体之间的复杂关联模式。随着这些粒子又与其他粒子碰撞并且发生纠缠，量子纠缠扩散出去，通过与环境的相互作用而遍布于整个空间，从而变得不可测量。由于这样的原因，光子之间的原始纠缠被擦去了。

然而，令人惊奇的是这些关联确实存在，在恰当调控的实验室条件下，人们可以在很远的距离下直接观测到这种关联性。这就告诉我们，从根本上来说，空间并不像我们所认为的那样。

那么时间又是怎样的呢？

# 第2部分
# 时间与经验

## 第5章　冰封之河

### 时间是流动的吗

在人们所接触过的各种概念中，时间是人们最熟悉但却最难以理解的一个。我们常说时光飞逝，我们也说时间就是金钱，我们总是试图节约时间，虚度光阴便感伤不已。但是，时间究竟是什么呢？按圣·奥古斯丁和波特·斯图尔特大法官[16]的说法，我们看一眼就知道时间是怎么回事。但是，在这新千年破晓之际，我们对时间的理解势必要深刻一些。事实上在某些方面，我们的确理解得深刻了一些。但在另一些方面，却不是这样。经过几个世纪的困惑和思考，我们已经洞悉到了时间的一些神秘之处，但留给我们的还有许多未解之谜。时间到底来自何方？一个没有时间的宇宙意味着什么呢？时间能像空间那样不只有一个维度吗？我们能够到过去“旅行”吗？如果能的话，我们可以改变某些事情的结局吗？时间有没有绝对意义上最小的量呢？时间是宇宙组成中真正的基本要素呢，还是单单为了协调人类感知而生的一种有用但却无法在写有宇宙的最基本原理的字典中找到的概念呢？时间是不是由某些尚未发现的更基本的概念派生出来的呢？

完备且令人信服的回答这些问题可算是当代科学家最雄心勃勃的目标。但科学家们要回答的并不仅仅是这些大问题。有些最棘手的宇宙学难题甚至来自于日常生活中的时间体验。

时间与体验

狭义相对论与广义相对论粉碎了时间的普适性和唯一性。根据相对论，我们每个人都拥有旧的牛顿体系中的普适时间的一块碎片。它成为我们个人的时钟，无情地把我们从一个时刻推到下一个时刻。相对论令我们震惊，因为当我们每个人的时钟滴滴答答地均匀地前进时，我们大家对时间的直觉感受没问题，但把我们的时钟与其他人的时钟相比时却会发现不同之处。你的时间没必要与我的时间一样。

我们可以把这种思想看做是一种给定条件。但对我而言，时间的真正本质究竟是什么呢？如果一开始就不与其他人的时间体验做比较，那么个人体验和构想的时间的全部特点是什么呢？这些体验有没有准确的反应时间之本性呢？关于实在性的本质，它们又会告诉我们什么呢？

我们的经验告诉我们，显而易见，过去不同于未来。未来代表了许多可能性，而过去则只有一种可能，就是实际发生的情形。我们有能力在一定程度上去影响、去塑造未来，而过去是不可改变的。在过去和未来之间的是现在的概念——每时每刻都在变化的短暂瞬间，就像电影中的画面，当放映机的强光扫过画面时就成为瞬间的现在。时间看起来以一种无休止的、完美又均匀的节奏不断前进，一次次的抵达每个一闪即逝的现在。

我们的体验也告诉我们时间具有很明显的方向性。比如我们没有必要为牛奶洒出而大惊小怪，因为一旦溢出来就不可能再回去了：我们从未见过洒出的牛奶自己汇聚起来，从地板上一跃而起，然后汇集到厨房柜台直立的玻璃杯里。我们的世界就像一支单向的时间之箭，从未偏离固定的模式：事物开始于此而终止于彼，但却不能反过来，开始于彼而终止于此。

因此，我们的经验告诉我们时间的两个特点。第一，时间看起来

是可以流动的。这就像我们站在时间之河的岸旁，看着汹涌澎湃的急流奔腾而去，每一朵未来的浪花经过我们的那一刻就成为现在，当急流远去奔向下游时就是过去。如果你觉得这种理解太过被动的话，可以把这个比喻颠倒一下：时间之河载着我们毫不停歇地向前驶去，从现在到下一刻，经过的景色远远褪去之时就成为过去，未来总在下游等待着我们（经验告诉我们，时间这个概念常常激发一些让人多愁善感的比喻）。第二，时间是有方向的。时间之流看起来朝向一个方向而且只能朝一个方向，这就意味着事情的发生只能有一种时间上的顺序。如果某人给你一盒牛奶溢出的胶卷，但胶卷被切割成了单独的几部分，通过查看这堆图像，你可以按正确的顺序重组这些图片，而完全不用胶片制作人给你任何指示或帮助。时间看起来有内在的方向性，从我们所谓的过去指向未来，事物总在变化——牛奶洒出，鸡蛋破碎，蜡烛燃烧，人会变老——普遍来说总是按照这个方向。

时间的这些最易于为人所感受的特点最使人困惑。时间真的会流动吗？如果答案是肯定的话，那么什么才是实际意义上的流动呢？时间这家伙流动得究竟有多快呢？时间真的有方向吗？举个例子来看，空间看起来就没有内在的方向——对于处在宇宙黑暗中的宇航员而言，左右、前后以及上下，都是一样的——那么时间的方向性是从何而来的呢？如果时间有方向的话，它是绝对的吗？或者说事情可以向时间之箭的反向演化吗？

让我们先在经典物理学的背景下，来看看我们对这些问题的理解。在本章其他部分和下一章（我们将会分别讨论时间的流动性和时间之箭）中我们将忽略量子概率和量子的不确定性。不过我们的讨论所得可以直接推广到量子领域，而在第7章中，我们就将从量子的角度来看看这个问题。

### 时间会流动吗

从有意识的人的角度来看，答案是显然的。当我打出这些字时，

清晰地感觉到了时间在流动。每一次按键，都意味着现在将让位于下一刻的到来。当你读这些字，当眼睛从一个字扫到下一个字时，你也一定感觉到了时间的流动。但是，虽然物理学家们努力尝试过，可没有人在物理定律中找到任何令人信服的证据，支持时间可以流动这种直观感受。实际上，对爱因斯坦狭义相对论思想的一些再思考却为时间不会流动提供了证据。

为了便于理解，我们来回忆一下第3章中介绍过的时空的面包片描述。面包条的每一切片是某个观测者的现在；每一片都代表着他或她眼中某一时刻的空间。这些切片一片接一片的按照观测者的体验排列起来的整体，就是一片时空区域。如果我们将这种设想推向极端，将每一片都想象成可以描述观测者眼中某一时刻的全部空间，如果我们再将从古老的过去到遥远的未来间所有可能的切片都考虑进来，这块面包就将代表所有时间内的整个宇宙——整个时空。每一个事件，无论何时何地发生，都可以用面包中的某个点来代表。

如图5.1所示，但这种描述法可能会令你抓狂。站在该图“外面”，我们可以看到整个宇宙，每一时刻的整个空间，这种图外人的角度是一种虚构的有利位置，没有人有过这种体验。我们都处在时空中。你或我曾经拥有的每一次体验都在某一时刻发生于空间的某个位置。因为图5.1描绘了整个时空，它包含了类似的所有体验——你的，我的，以及每个人和每一件事情。如果你能把镜头推近并密切关注地球上所发生的一切，你将会看到亚历山大大帝正在上亚里士多德的课，列奥纳多·达·芬奇在为蒙娜丽莎画上最后的一笔，乔治·华盛顿横渡特拉华河；[17]你从左到右继续观看，就将看到你的祖母正在跟一个小女孩玩，你父亲在庆祝他的第10个生日以及你在学校的第一天；再往右边远一点的图像看去，你会看到自己正在看这本书，你曾孙女出生了，再远一点，有她成为总统的就职典礼。图5.1的分辨率太过粗糙，实际上你不会看到这些，但你能看到太阳和地球的构造史（图解），从它们诞生于气体凝合到太阳变成了红巨星时的地球灭亡。所

有发生的事情都可以看到。

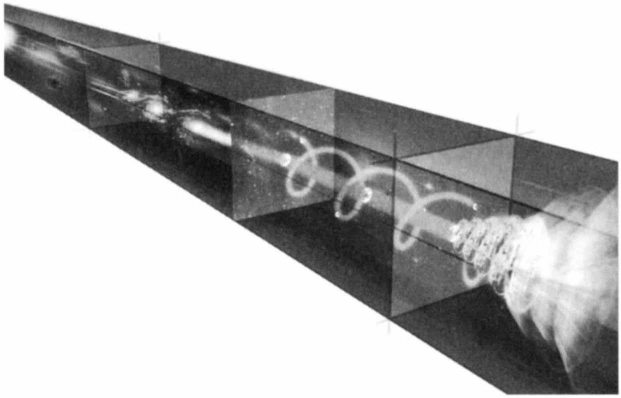


图5.1　所有时间中的全部空间的示意图（当然，图上画的只是一段时间中的部分空间）。图中画出了某些早期星系的形成，太阳和地球的形成，还画出了当太阳终于燃烧殆尽最终成为红巨星时地球的终极命运。我们的未来就在这张图中。

毫无疑问，图5.1是想象出来的，它位于空间和时间之外，没有哪个地方也没有哪个时刻能提供这样的视角。虽然如此——虽然我们实际上无法摆脱时空的限制，遍览宇宙的全貌——图5.1的描述还是为我们提供了一种分析和弄清楚空间和时间基本特性的有力方法。作为主要的例子，在这一框架下，时间流动性的直观感受可以用电影放映机比喻的变体生动地勾画出来。想象有一束光，一片接一片的照亮时间片，使每一时间片短暂地亮一下——使时间片成为瞬间的现在——当光照射到下一个时间片时作为现在的时间片就即刻熄灭。现在，按照这种直观方式思考时间，光照亮了某一切片，而时间片中的你在地球上，正在读这些字；光又照亮了另一切片，而另一时间片中的你还在地球上，正在读这些字。但是，又一次，虽然这种图像看起来与日常经验相一致，科学家们却无法找到适合的物理原理来描述这样一种活动的光。他们仍未找到这样一种物理机制，当其朝着未来不断演化时，能够使某一时刻瞬间变得真实——变成瞬间的现在。

正相反。尽管图5.1的视角是想象出来的，但却有令人信服的证

据表明，时空条——整个时空，而不是单个的时空片——是真实的。爱因斯坦的工作中尚未引起普遍重视的一点是，在狭义相对论中，所有的时刻都具有同等的地位。虽然现在的概念在我们的世界观中起着重要的作用，但相对性却要再一次颠覆我们的直觉，它声称我们的宇宙是一个平等的宇宙，每一时刻都是同样真实的。第3章中在狭义相对论的框架下讨论旋转的桶的问题时，我们就曾遇到过这个问题。在那里，通过类似于牛顿式的间接推理，我们得出结论，时空足可以作为加速运动的基准。在这里，我们从另一个角度再来考虑这个问题并进一步深入。我们认为图5.1中的时空条的每一部分与其他部分具有同等地位，这正表明，就像爱因斯坦所相信的那样，过去、现在和未来具有同样的实在性，我们所想象出来的时间之流——时空片一片接一片的变得光亮或黯淡——只是一种幻觉。

过去、现在和未来的持续幻象

为了便于理解爱因斯坦的观点，我们需要实在性的有效定义，如你愿意的话叫做算法也行，以便明确某一给定时刻都存在着哪些事情。现在给出一种通用的办法。当我考虑实在性——在这一时刻存在哪些东西——时，我在头脑中立刻勾画出了一幅快照，此时此刻整个宇宙的静止图像。当我打下这些字时，我对此时此刻存在什么的感觉，对实在性的感觉，可以列很长一张目录——午夜时分厨房时钟的滴答声；我家的猫在地板和窗沿之间攀爬；照亮都柏林清晨的第一缕阳光；东京股票交易所的喧闹声；太阳中两个特殊氢原子的融合；猎户座星云所发射出的光子；垂死的恒星衰变为黑洞的最后一刻——这些就是此刻我头脑中所出现的静止图像。这些就是此时此刻正在发生的事情，因此，它们就是我所宣称的存在于此刻的事物。查理曼大帝现在还在吗？不。尼禄现在还在吗？不。林肯现在还在吗？不。埃尔维斯[18]现在还在吗？不。他们当中没有一个出现在我现在的目录中。现在有人在2300年或3500年或57000年出生吗？不。他们中没有一个出现在我头脑中的静止画面里，没有一个在我现在的时间片中，因

此，也没有任何一个在我目前的现在列表中。因此，我毫不犹豫地说，他们现在不存在。我就是这样定义任一给定时刻的实在性；这是我们当中大多数人思考存在性时，虽然常常是不知不觉中，但常用的一种直观的方法。

在下面的讨论中，我将会用到这样的概念，但仍然要警醒棘手的一点。一张关于现在的目录——用这种方法来思考实在性——是一件很有意思的东西。你此刻所看见的一切事物都不会出现在你的现在的目录里，因为光需要花一段时间才能到达你的眼睛。任何你看到的事情都是已经发生过的了。你现在读到的该页中的文字这事并不是现在发生的；实际上，如果书离你有1英尺（1英尺≈0.3048米）远，你所看到的字是它们十亿分之一秒之前的样子。如果你在房间中四处看看，你所看到的一切都是它们十亿分之一秒或二十亿分之一秒之前的样子；如果你的目光贯穿整个大峡谷[19]，你所看到的是它万分之一秒之前的样子；当你看月亮时，你看到的是它一秒半之前的情形；当你看太阳时，你看到的是它8分钟之前的情形；对于裸眼可见的恒星而言，你看到的是几十年乃至1万年之前的情形。令人惊奇的是，虽然头脑中的静止图像描述了我们对于实在性的感觉，我们对“那儿有什么”的直观感觉，但它所包括的却是我们此刻不能去体验，或者影响，甚至不能现在就记录的事件。事实上，一张现在的目录只能事后编辑。如果你知道某物距离你有多远，你就能决定现在所看到的光是何时发出的，因而你就能决定它到底属于哪个时间片——上面应当记录着已经过去的时刻的现在目录。不过，这点正是关键，当我们用这些信息去编辑任意给定时刻的现在目录时，我们得根据从更远的源头收集到的光信号不断更新这张目录，上面记录的事情正是我们直觉上相信发生于那一刻的事情。

奇怪的是，这种直截了当的思考方式将会出人意料地扩展实在性的概念。你想，根据牛顿的绝对空间和绝对时间概念，在任一个给定时刻，每个人头脑中的宇宙静态画面都应该包含相同的东西；每个人

的现在都是同样的现在，因此所有人在某一时刻的现在目录都是一样的。如果某人或某物在你的某一时刻的现在目录上，那它必然也在我的同一刻的现在目录上。大多数人的直觉仍然是这种思维方式，但是相对论却告诉我们不应当如此。再看一下图3.4。处于相对运动中的两个观测者都有现在——从每一个人的角度来看，都只是某个时间点——但两者的现在却是不同的：两者在时空中按不同的角度切割他们各自的现在时间片。不同的现在意味着不同的现在目录。相对于彼此运动的观测者对于某一时刻存在什么有不同的概念，因而他们对于实在性有不同的概念。

在日常生活的速度水平下，两个观测者的现在时间片之间的角度差异是十分微小的；而这就是为什么我们在日常生活中感受不到我们所定义的现在和别人所定义的现在有什么区别。由于这个原因，大多数狭义相对论的探讨都集中在如果我们以非常大的速度——接近于光速的速度——运动时将会发生什么上，因为这样的运动将会显著地放大相对论的效应。不过，将两个观测者对现在的定义之间的差别放大，还有另外一种方法，在我看来，这种方法会对解决关于实在性的问题有独到的启迪。这种方法建立在下列的简单事实上：假如你我以略微不同的角度切开一块普通的面包，则剩下的面包片将不会受到多大影响。但如果面包非常巨大，结果就全然不同了。就像一把巨大的剪刀，只要稍稍张开一点，它所展现的刀锋就将极其巨大；要是面包条足够巨大的话，两个切片的角度只要差一点点，它们彼此之间的差别就将极其巨大。参见图5.2。

对于时空而言也是一样的。在日常速度下，对于处于相对运动状态的两个观测者而言，描述现在的时间片的方向之间只有一个微小的角度。如果两个观测者距离很近，几乎不会产生什么影响。但是，就像长条面包一样，即便角度很小，可如果要探讨的是非常大的距离的话，切片之间也会产生巨大的差距。对于时空片而言，不同片之间的巨大偏离就意味着不同观测者对现在发生的事件的认识存在着巨大的

差异。如图5.3和图5.4所示，这就意味着相对于彼此运动的个人，即使是以普通的日常速度运动，但只要空间上相隔很远，也会有不同的现在概念。

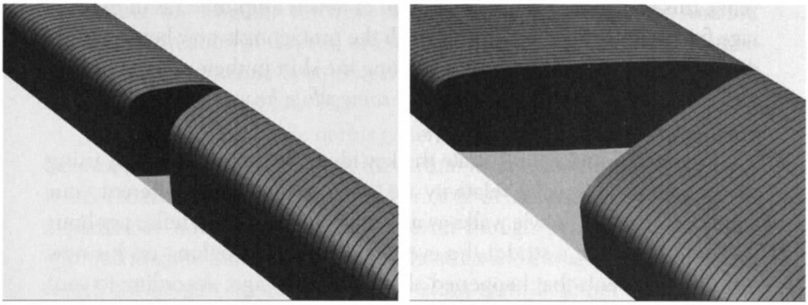



图5.2　（a）：以略微不同的角度切开一块普通的面包，切片之间将不会分离多少。（b）：对于大面包而言就不一样了，虽然还是以相同的角度切开，但面包越大，切片之间的偏离就越大。

为了使讨论更加具体，想象一下丘巴卡[20]。他在一个非常非常遥远的行星上——距离地球大概有100亿光年——他正懒散地坐在他的卧室里。再进一步假设你（只是静静地坐着在读这本书）和丘巴卡相对于彼此静止（简单起见，忽略行星的运动、宇宙的膨胀、引力效应，等等）。由于你和丘巴卡相对于彼此静止，因此在时间和空间问题上，你们将达成一致：你们两人将以类似的方式切割时空条，也就是说你们的现在目录将会彼此吻合。过一小会儿，丘巴卡站起来去散步——非常放松的漫步——但朝着远离你的方向。丘巴卡运动状态的变化意味着他的现在概念，他的时空切片，都将发生轻微的旋转（参见图5.3）。这种角度上的微小变化在丘巴卡附近不会产生什么明显的效应：他新定义的现在概念，同在他的卧室里的其他人的现在概念之间的差异非常小。但是如果相距100亿光年的话，丘巴卡的现在概念上的这种微小变化将会被放大[如图5.3（a）和图5.3（b）所述，但是如果所要讨论的两个点距离很远，则这两个点现在的微小改变将被清楚地放大]。虽然在丘巴卡静坐时，他的现在和你的现在是一样

的，但由于丘巴卡的轻度运动，你们两人的现在变得完全不一样了。

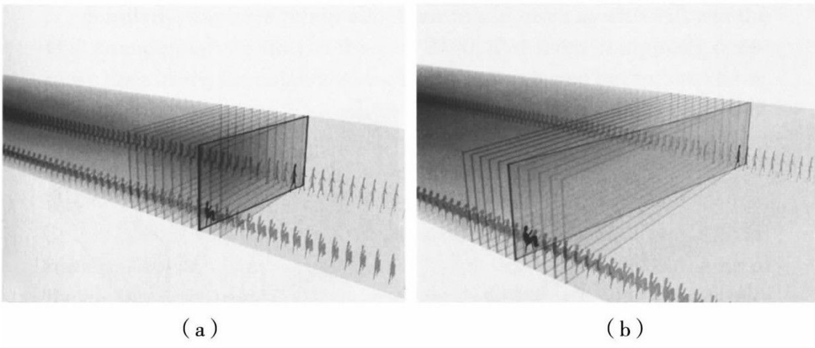


图5.3　（a）：两个相对于彼此静止的人对于现在有相同的概念，因此就会有相同的时间片。如果一个观察者远离他们的时间片——每个观察者眼中的现在——相对于彼此则发生了旋转；如图所示，对于运动的观察者而言，变黑的现在的时间片旋转到静止的观察者的过去的时间片中。（b）：观察者之间偏离得越远，时间片产生的偏离就越大——他们对于现在的概念偏离就越大。

图5.3和图5.4用图示的方法阐释了关键思想，但运用狭义相对论的方程，我们可以计算出你们的现在差别到底有多大。[1]如果丘巴卡以每小时10英里的速度远离你而去（是的，丘巴卡大步流星地走着），那么在他的新的现在目录里地球上所发生的事情，对你而言，其实是150年前发生的。依照他的现在概念——他的概念与你的概念同样有效，并且就在刚才你们俩对于现在的概念还完全一致——你还没有出生。如果他以相同的速度朝你走来，如图5.4所示的那样，角度变化的方向相反，那么他所谓的现在对你而言，将是未来150年后！这样看来，按照他所谓的现在，你不再是这个世界的一部分。假如，丘巴卡不是走，而是跳进了千年帝国之鹰飞船[21]以每小时1000英里的速度（比协和式超音速客机[22]的速度慢一点）飞行，如果他的方向是离你而去，那么他所谓的现在对你而言，将是15000年前地球上发生的事情，反之则是未来15000年后所发生的事情。如果方向、运动速度合适的话，猫王、尼禄、查理曼大帝、林肯或某个未来才出生的人，都有可能出现在他的现在目录上。

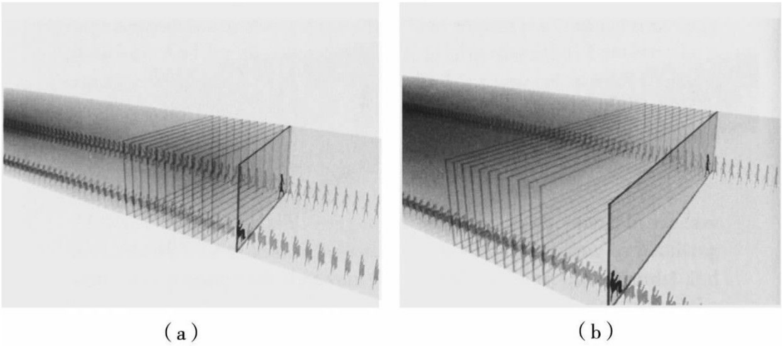


图5.4　（a）和图5.3（a）的唯一区别在于，当一个观察者朝另一个观察者运动时，她的现在片会转到另一个观测者的未来，而非过去。（b）同图5.3（a）一样——在同样的相对速度下，更大的间隔意味着现在概念上更大的分歧——只不过转动指向未来而不是过去。

尽管令人惊讶，但却不会产生任何矛盾，就像我们前面所解释的，某物距离你越远，接受它所发散出的光就需要越长的时间，从而决定它应该属于哪个现在目录也需要花更长的时间。举个例子来说，即便正在前往福特大剧院总统包厢的约翰·维尔克斯·布思[23]在丘巴卡新的现在目录上（此时丘巴卡正站起来，以每小时9.3英里的速度远离地球而去[2]），丘巴卡也无法采取任何行动来拯救总统林肯。这么遥远的距离，将需要许多时间来接收和交换信息，因此，实际上只有丘巴卡几十亿年后的后裔，才会接收到有关那一夜的华盛顿的光。问题在于，当他的后裔用这个信息来更新过去的现在目录时，他们将会发现林肯的暗杀与丘巴卡站起来远离地球而去都在相同的现在目录上。而且，他们也将发现在丘巴卡站起来前一瞬间，他的现在目录也包含了21世纪的你正坐在那儿读这段话。[3]

类似的，有一些关于未来的事情，比如谁将赢得2100年的美国总统大选，看起来是完全开放的：此次竞选的候选人很有可能还没有出生，更不用说决定竞选了。但是如果丘巴卡从椅子上站起来以每小时6.4英里的速度朝地球走来，他的现在目录——他对于现在存在什么，发生了什么的认识——将包括22世纪第一届总统的选举。对于我们而言还未决定的一些事情，在他看来却已经发生了。又一次，丘巴

卡再过很多亿年才能知道选举结果，因为得花那么长时间，我们才能把信号传递给他。但是当丘巴卡的后裔收到选举结果用来更新丘巴卡的历史册页，更新他过去的现在目录时，他们发现选举结果居然和丘巴卡站起来开始走向地球的时刻记录在同一张现在目录上，丘巴卡的后裔注意到，比这张现在目录早了一点点的现在目录中，记录着你在21世纪的某一天看完这段文字的事件。

这个例子有两点非常重要。第一，虽然对于接近光速时相对论效应会变得非常明显这一事实，我们已经习以为常；但还需要知道，对于低速运动，如果空间上能够相距很远，那么相对论效应也会得以放大。第二，这个例子对下面的问题很有启发性，即时空（面包条）究竟是真的实体还是只是一种抽象的概念，一种空间的现在和它的历史以及所谓未来所组成的抽象整体。

你看，丘巴卡关于实在性的观点，他头脑中定格的画面，他对于现在存在何物的概念与我们的实在性观念是一样真实的。因此，在评价实在性的构成时，如果我们不考虑他的观点，那就未免太狭隘了。对于牛顿而言，这样一种平等主义的做法并不会有多大的不同，因为，在一个有绝对空间和绝对时间的宇宙里，所有人的现在时间片都是一致的。但在相对论的宇宙里，也就是我们的宇宙里，这样的平等主义就会带来很大的不同。尽管我们熟悉的关于现在存在何物的概念只相当于单独的一片现在时间片——我们通常把过去看做已经逝去的，未来是还没有发生的——我们却不得不将丘巴卡的现在切片一起考虑来扩大我们的认识，就如上文中所讨论的，他的现在切片与我们的有很大的不同。此外，由于丘巴卡最初的位置和他移动的速度是任意的，我们必须得将与所有可能性有关的现在时间片都包括进来。这些现在时间片，正如我们上文中所讨论的，将以丘巴卡——或者其他或真实或假设的观测者——在空间中的初始位置为中心，根据给定速度的不同而旋转一定的角度。（唯一的限制是光速的限制，在尾注中将进一步解释，根据图示，光速的限制相当于旋转角度最大为45度，

顺时针或逆时针均可。）正如你在图5.5中所看到的，所有的现在时间片充斥于整个时空条。事实上，如果空间是无限的——如果现在时间片能向无限远处扩展——那么旋转的现在时间片可以任意远为中心，因此它们的集合可以遍布于时空中的每一点。[24]

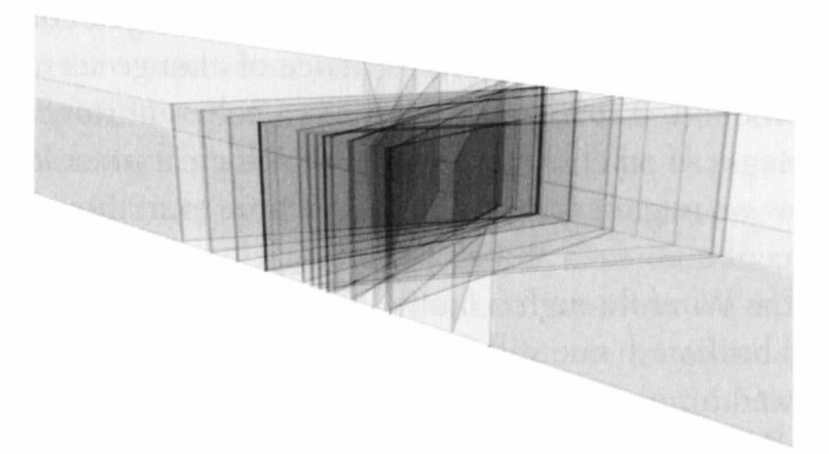


图5.5　一个不同观测者（不管是真实的还是假想中的）的现在片的例子，这些观测者距离地球不同远近，速度也各异。

因此，如果你认为实在性由你现在头脑中定格的事情组成，如果你同意你的现在概念与位于远方空间中可以自由移动的某人的现在概念一样有效，那么实在性将涵盖时空中所有事件。整个面包都存在。就像我们视所有空间都真实的存在一样，我们也把所有时间（包括过去、现在和未来）视为真实的存在。过去、现在和未来显然是有区别的。但是，就像爱因斯坦曾经说的“对于我们这些充满信心的物理学家而言，过去、现在和未来之间的区别只是一种幻觉，虽然它总是存在的”。[5]唯一真实的事物就是整个时空。

体验和时间的流动

以这种方式来思考问题的话，虽然从不同的视角来看事件发生的时间不同，但它们总是存在的。它们永远占据了时空中的某一点。它

们并没有流动。如果你在1999年新年除夕的午夜度过了非常愉快的时光，你一直都会拥有它们，因为那是时空中不可变的一个点。接受这种说法有些困难，因为我们的世界观对过去、未来和现在有着截然不同的看法。如果我们固执于熟悉的时间观念，就会发现它将在现代物理冷酷的事实面前碰壁，它所能有的唯一的安身之处就是人类的意识。

不可否认，我们的意识体验似乎遍布于整个时空切片。打个比方来说，我们的思想就好比先前提到过的放映机的光，当时间的某时刻被意识的力量照亮时，它们就成为鲜活的画面了。从某一刻到下一刻的流动感源于我们的思想、感觉和认知在意识上的改变。改变的结果将会导致持续的运动；它会发展成前后一致的故事。但是——不依靠任何心理学或神经生物学的借口——我们可以想象一下我们是如何感受时间的流动，即使实际上并没有这样的事情发生也没关系。为了便于理解我的意思，想象一下现在有一台有点毛病的DVD播放器，它会随意的前进或后退，我们用它来播放电影《飘》：屏幕上刚才还放映某一刻的画面，但下一刻立刻就切换成了完全无关的画面。当你观看这种跳跃性的画面时，你可能很难弄清楚到底在演什么。但对于郝思嘉和白瑞德来说没有问题。在每一帧画面中，他们做他们在那一帧画面中总会做的事情。如果你把DVD停在某个特殊的画面，问他们相关的想法和记忆，他们给你的答案将与DVD功能正常时他们会给你的答案一模一样。如果你问他们是否因南北战争的混乱顺序而迷惑，他们将会疑惑地看着你，认为你一定是喝了太多的冰镇薄荷酒。在任意给定的画面里，他们将会有画面那一刻的思想和回忆——特别是，那些想法和记忆给他们的感觉是时间像平常一样平稳而连贯地逝去。

类似的，时空中的每个时刻——每个时间片——就好比一部电影中的某一帧静止画面。画面的存在与否取决于是否有光照亮它。就像郝斯嘉和白瑞德一样，对于正处于任何这样时刻的你来说，这就是现在，“现在”就是你当时感受到的那一刻。并且“现在”永远都是你正在

感受到的那一刻。而且，在每一个独立的时间片里，你的思想和记忆都足以使你产生时间在不断地流向下一刻的感觉。这种感觉，这种时间正在流动的意识并不需要之前的时刻——之前的画面——来“连续放映”。[6]

稍稍想一下，你就会意识到这是一件非常好的事情，由于另外的更为基本层面的原因，放映机所发出的光有序地将时间激活的概念有非常严重的问题。假如放映机正常地放映着某一瞬间的画面——比如说1999年新年夜的午夜敲钟场面——突然画面暗了下来，意味着什么呢？如果某一时刻已被点亮，那么处于照亮状态就是那一时刻的特性之一，该特性也应该像发生在那一时刻的其他事情一样永恒而无变化。历经照亮——“活”起来，成为此时，成为现在——然后再回归黑暗——“休眠”，变成过去——就是经历变化。但变化的概念与单独的时刻无关。变化将不得不通过时间来发生，变化标志着时间的流逝，但时间的概念究竟是什么呢？从定义上看，时刻并不包括时间的流逝——至少不是我们所说的时间——因为时刻是时间的原材料，并不会变化。某个特殊时刻不再变化就像空间中某个特殊位置一样：如果某位置变化，它就是空间中的另一个位置了；同理，如果某时刻变化，它就是另一个时刻了。放映机的光激活每一个新的现在这样的直观图像经不起仔细的推敲。换句话说，每一时刻都被照亮，每一时刻都会保持其被照亮的状态。每一时刻都是这样。仔细想来，时间的河流更像是一块巨大的冰块，每一时刻都永远的冰冻在它自己的位置上。[7]

这样的时间概念与我们的内在感受非常不同。虽然这种概念源于爱因斯坦的洞察力，可他本人也很难完全接受这种观念上的深刻转变。鲁道夫·卡那夫[8]叙述了他和爱因斯坦之间就这个问题展开的精彩对话：“爱因斯坦说有关现在的问题困扰着他。他解释说有关现在的体验对人类来说意味着某种特殊的东西，一种从本质上不同于过去和未来的东西，但这种重要的不同却不会也不能出现在物理中。这种无法被科学理解的体验似乎让他很头疼但却不得不顺从。”

这种顺从就给我们提出了一个关键问题：究竟是科学不能像解释肺可以吸入空气那样，轻易地解释存在于人们意识中的时间的基本特性呢？还是人类意识强加给时间一种人为的特性，因而无法用物理定律来解释呢？如果你在工作日问我这个问题，我将赞成后一种观点，但夜幕降临，当重要的思想都变为日常生活惯例时，就很难完全抵制前一种观点了。时间是一门深奥的科目，我们还远远没有理解它。很可能未来的某一天，某个聪明的人发现了一种新的看待时间的方式，揭示了流动的时间的真正物理学基础。以上建立在逻辑和相对性基础上的讨论，可能就是故事的全部。当然，时间流动的感觉在我们的生活体验里根深蒂固，并且遍布于我们的思想和语言中。我们已经而且将继续误入用习惯性的口语描述时间的流动这样的歧途。但不要把语言和实在性搞混淆了。比起深刻的物理定律，人类语言更善于描述人们的体验。

## 第6章　偶然和箭头

### 时间有方向吗

即使时间并不流动，探究时间是否有方向——事物在时间中的发展演变是否有一个可以用物理原理来辨认的方向——仍然自有其意义。这个问题等于是在问，事件在时空中的分布是否存在某种固有的顺序？事件按时间顺序发生与逆着时间顺序发生会有什么不同？就像我们每个人所知道的那样，两者之间一定存在着巨大的不同；正是由于这种不同，生活才会既充满希望，又令人痛苦不堪。但是，我们将会看到，解释过去和未来之间的不同之处比你想象的还要困难。而更令人惊讶的则是，我们将要解答的问题与宇宙起源时的具体条件有着密切的联系。

### 谜团

每一天中，我们都有成百上千次的机会看出顺着时间方向发生的

事件和逆着时间方向发生的事件之间的巨大区别。滚烫的比萨从烤箱中拿出的过程中会冷却下来，但我们从未看到过比萨从烤箱中拿出后会变得比以前更热。放进咖啡中的奶油搅匀后会变成均质的棕褐色液体，但我们却从未看到一杯淡咖啡不经搅拌，自己会分离出白色奶油和黑色咖啡。鸡蛋坠落、打碎并破碎出来，但我们却从未看到破碎的鸡蛋和鸡蛋壳自己聚集起来，形成未破碎的鸡蛋。当我们拧开可乐瓶时，压缩的二氧化碳气体会跑出来，但我们却从未看见过分散的二氧化碳聚集起来并嗖的一声返回瓶中。室温环境中的杯子里的冰块会融化，但我们却从未看到杯子里的水珠会在室温下凝结成冰。这些习以为常的事件，连同数不胜数的其他事件，只沿一个时间方向发生。它们从不会逆着时间方向发生，因此它们为我们带来了先和后的概念——它们给我们带来了稳定可靠且具有普适性的过去和未来的概念。这些现象使我们确信，从外部（如图5.1所示）观测整个时空的话，我们将看到时间轴具有明显的不对称性。鸡蛋已经破碎的那个世界在时间轴的一端——传统上我们将其称之为未来——而对应着的另一端就是鸡蛋尚未破碎的世界。

或许最显而易见的例子是，我们的意识可以存储被我们称之为过去的许多事情——这就是所谓的记忆——但却没人能够记住被我们称之为未来的事情。因此，很显然，过去和未来之间存在着很大的不同。各种各样的事情在时间的长河中总是沿着确定的方向发生。我们能回忆起的事情（过去）和不能回忆起的事情（未来）之间有着明显的区别。这就是为什么我们会说时间具有方向性或有一个箭头。[1]

物理学和广义上的科学，都以规律为基础。科学家们研究自然，发现规律，并用自然定律来解密这些规律。因而，你可能会认为，使我们清楚地感受到时间之箭的各种各样难以计数的规律性，意味着存在这样一条基本的自然定律。构建这样一条定律最笨的办法就是引进溢出牛奶定律或破碎鸡蛋定律，前者说的是从牛奶溢出来就不会自己再汇聚起来，后者则是鸡蛋破碎出来就不可能再自己聚集起来形成一

个完整的鸡蛋。但这样的定律对我们毫无用处：它只是描述性的，只是简单的说明观测到发生了什么，而无法提供任何解释。但我们期盼着物理学最深奥的领域中存在着某种不这么傻的定律，我们可以用它来描述组成比萨、牛奶、鸡蛋、咖啡、人和星球的粒子——组成一切事物的基本成分——的运动和性质，这个定律将会告诉我们事物为什么会按照某种特定的顺序演化而不能反过来。该定律将给予我们所观测到的时间之箭一个基本解释。

但令人头疼的是没有人发现这样的定律。而且，从牛顿到麦克斯韦，到爱因斯坦，他们所发现的物理定律，以及今天的所有物理定律，都显示出过去和未来之间存在着一种完美的对称性。[25]我们并未在这些定律中发现只可沿着时间轴的某个方向应用该定律的限制条款。这些定律应用于时间轴的不同方向时不会有什么区别。过去和未来在这些定律下看来都是一样的。即使我们的经验一次又一次告诉我们，事件如何随时间发展存在一定的方向性，但这样的时间之箭却不存在于基本的物理定律中。

## 过去、未来和基本物理定律

怎么会这样？物理定律没有提供用以区分过去和未来的基础吗？为什么会没有物理定律能够解释事件只能按这种顺序发展而不能逆过来呢？

这种情况更为令人迷惑。众所周知的物理定律实际上声明——与我们的生活经验相反——奶油咖啡可以分离成黑色的咖啡和白色的奶油；破碎的蛋黄和破碎的蛋壳能自己聚集起来形成一个完美光滑的鸡蛋；室温下水杯中融化的冰水可以重新形成冰；你打开苏打水时放出来的气体可以自己返回瓶中。我们现今所知的所有物理定律都完全支持所谓的时间反演对称性。这种对称性说的是，如果事件可以按照某种时间顺序发展（奶油和咖啡混合，鸡蛋打碎，气体溢出），那么这些事件也可以按照相反的方向发展（奶油和咖啡分离，鸡蛋完好如

初，溢出的气体回到瓶子里）。简短地用一句话来总结就是，物理定律不仅没有告诉我们事件只能按某种方向发展，而且还从理论上告诉我们事件可以向相反的方向发展。[26]

但重要的问题是，为什么我们从没有看到这样的事情发生呢？我敢打赌一定没有人亲眼看见打碎的鸡蛋聚集起来恢复成原样。但是如果物理定律允许这种情况存在，而且这些定律平等地对待打碎的鸡蛋和未打碎的鸡蛋的话，为什么一种情况从未发生而另一种情况总是发生呢？

时间反演对称性

解决上述谜团的第一步需要我们更为扎实的理解已知物理定律为什么满足时间反演对称性。为了这个目的，这样想象一下：现在是25世纪，你与你的搭档库斯托克·威廉姆斯在新的星际联盟打网球。由于不太习惯金星上较低的引力，库斯托克用力过猛，一个反手球将球打到了深不可测的漆黑星空中。一架正在经过的太空飞船拍摄到了飞驰而过的球，并把胶片送到了CNN（星际新闻网）[27]播出。这儿有一个问题：如果CNN的技术人员犯了错误，把这段网球的片段反过来放映，人们是否能看出来呢？如果你知道拍摄时摄像机的朝向，你可能会指出他们的错误。但是，如果没有任何其他信息，只看底片的话，你能挑出他们的错误吗？答案是否定的。如果顺着时间方向，底片将显示球从右飞向左，如果反过来就会变成球从左飞向右。当然，从经典物理学定律的角度来看，网球朝左或朝右运动都是可以的。因此无论片子是顺着时间的方向还是逆着时间的方向放映，你所看到的运动与物理定律完全一致。

上文中，我们一直在使用这样一个假设，即，没有力作用于网球，因而网球是匀速运动的。现在我们把力加进去考虑一些更普遍的情况。根据牛顿定律，力的作用将会改变物体的速度：力意味着加速度。做了上述的假设后，我们再来看看网球的情况：网球在空中飞行

时，由于受到木星引力的作用，向下加速运动，朝着木星表面向右划出一段美丽的弧线，如图6.1（a）和图6.1（b）所示。如果你逆着放这段运动的底片的话，网球将向上加速运动，因而会朝远离木星的方向划出一道弧线，如图6.1（c）所示。现在有了个新问题：底片所描述的逆向打网球的运动——实际上所拍摄到的运动的时间反演运动——是经典物理学定律所允许的吗？这种运动会在真实的世界中发生吗？乍看之下，答案显然是肯定的：网球的运动轨迹既可以是向右下的弧线，又可以是向左上的弧线，或者是数不清的其他轨迹。那么，困难之处到底在哪？这个嘛，虽然答案确实是肯定的，但推演过于草率而忽略了问题的真正内涵。

当逆向放映片子时，你会看到当网球撞击到木星时，会以相同的速度（但以完全相反的方向）远离木星，朝左上的方向运动。片子的最初部分显然与物理定律相一致：举个例子来说，我们想象一下，某人在木星表面以该速度击出网球就符合这种情形。关键的问题在于，逆向运动的其余部分是否与物理定律相一致。以该初速度击出的网球——在受到木星向下的引力的作用下——实际上会沿着片子其余部分中所描述的逆向运动的轨迹运动吗？运动反过来之后，它会顺着原始的向下的轨迹运动吗？

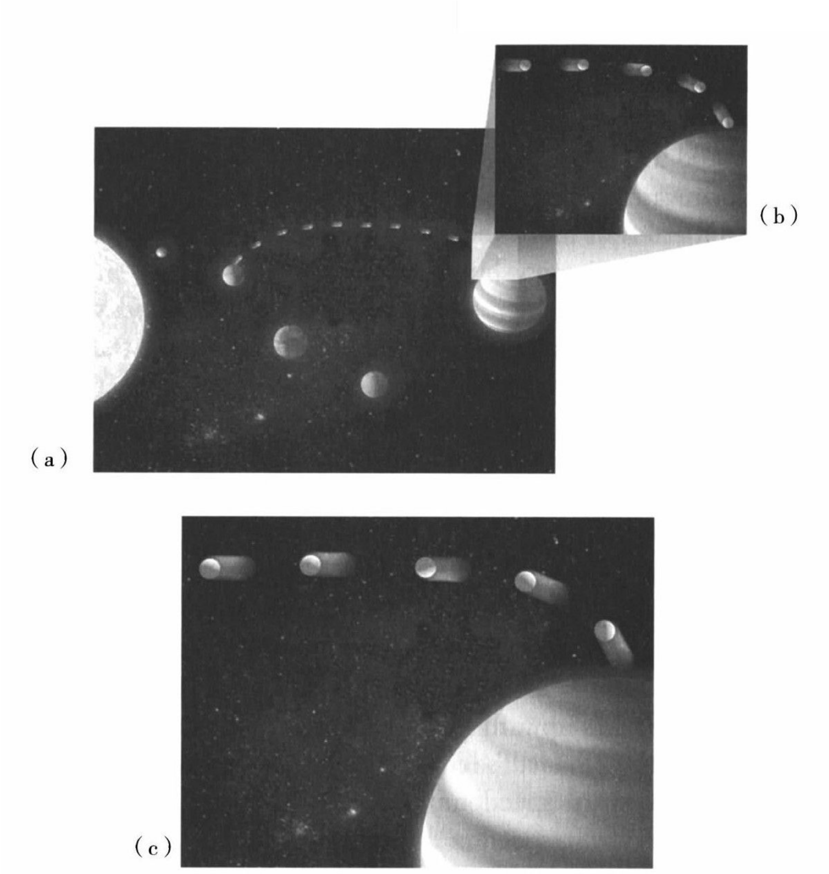



图6.1　（a）：从金星飞到木星的一个网球。（b）：特写。（C）：在撞到木星之前，网球的速度反向，新的运动轨迹。

这些更为精炼的问题的答案是肯定的。为了避免混淆，我们先把它讲清楚。图6.1（a）中，在木星的引力产生有效作用之前，网球是纯向右运动的。图6.1（b）中木星的引力有效地作用于网球，产生一个将它拉向地心的力——正如你在图中所看到的，引力的方向大部分是竖直向下的，不过也有一部分是向右的。这就意味着，当网球接近于木星表面时，它向右的速度将会略微有所增加，而它向下的速度将会增加得非常多。因此，在逆向放映的片子中，从木星表面击出的网球会略微向左主要向上运动，如图6.1（c）所示。木星的引力将对网

球向上的速度产生重要影响，使它越来越慢，同时也会减慢球向左的速度，只是没有那么夸张而已。随着球向上的速度的迅速减少，它速度的方向将主要向左，进而使得向上的弧线的运动轨迹偏左。接近弧线的末端，网球向上的速度和在下降过程中因木星引力而产生的额外的向右的分速度，将在引力作用下变为零，此时球以原始大小的速度向左运动。

上述论证都可以定量研究，但值得注意的关键之处在于，该运动轨迹恰与网球初始的运动轨迹相反。如图6.1（c）所示，简单的逆转球的速度——速度相等，但方向完全相反——我们可以使它完全沿着原来的轨迹运动，只是方向相反而已。我们再回到片子的讨论上，我们所看到的向上偏左的弧形轨迹——我们用以计算轨迹的是牛顿的运动定律——正是我们将片子逆过来放映所看到的。因此，逆向放映的片子所描述的网球的时间反演运动，和时间上正向运动一样，都遵守物理学定律。逆向放映电影时我们所看到的运动，在真实世界中可以实际发生。

虽然上述讨论中有一些细节被我放到了注释里，但其结论仍然具有普适性。[2]所有已知和广为接受的有关运动的定律——从刚才讨论过的牛顿的经典力学，到麦克斯韦的电磁理论，再到爱因斯坦的狭义和广义相对论（记住，我们将在下一章讨论量子力学）——都具有时间反演对称性：按时间轴正向发生的运动同样也可以逆着时间轴发生。由于术语有点混乱，我再来强调一下，不是把时间反过来。时间仍然保持原样。我们的结论是，要想使一个物体的运动轨迹逆转，只要在其路径上任意一点逆转其速度即可。同样的，相同的程序——在其路径上任意一点逆转其速度——将使物体按我们在反向放映的片子中所看到的方式运动。

网球和破碎的鸡蛋

观察网球在金星和木星之间运动——无论朝向哪个方向——并非

十分有趣。但既然我们所得出的结论可以广泛应用，我们现在就来看一些更加有趣的地方吧，比如说你的厨房。把一个鸡蛋放在厨房的餐桌上，让它沿着桌边滚动，然后掉到地上摔碎。可以肯定的是，在这一系列事件中存在许多运动。鸡蛋掉下来，蛋壳摔碎了，蛋黄溅得到处都是，地板震颤，周围的空气中形成漩涡。摩擦产生热量，使鸡蛋、地板以及空气中的原子和分子运动得更快。但是，正如物理定律告诉我们的，如何才能使网球丝毫不差地逆着原来的轨迹运动；同样的定律也会告诉我们如何才能使每一片蛋壳碎片，每一滴蛋黄，每一块地板，每一团空气精确的逆着原来的轨迹运动。我们所需做的“全部”只是将碎鸡蛋每一块碎片的速度反过来。更准确地说，我们在网球问题上的分析告诉我们，只要我们能把与鸡蛋破碎直接或间接相关的每一个分子和原子的速度都同时逆转过来，那么整个鸡蛋破碎的运动就会反过来进行。

再强调一次，就像网球运动一样，如果我们能成功逆转所有的速度，我们所看到的就像一部反向放映的电影。但是，不同于网球之处在于，鸡蛋破碎的逆运动将给人留下极其深刻的印象。厨房各处空气分子碰撞和微小的地板震动所产生的波汇集在碰撞的位置，造成每一片碎蛋壳和每一滴蛋黄都朝着发生碰撞的位置运动。每一种成分都以最初鸡蛋破碎过程中的速度运动，只是方向都相反而已。无数滴蛋黄都飞回形成一个球，就像无数片碎蛋壳完美的排列在一起形成一个光滑的卵形容器。空气和地板的震动与结合在一起的蛋黄和蛋壳碎片的运动配合得非常完美，形成一个重新组合的鸡蛋，恰好反弹离开地板，向上飞到厨房的餐桌上，轻巧地落在餐桌边缘，然后滚动几英寸（1英寸≈2.54厘米）优雅地回到远处。如果我们逆转全程中每一样东西的速度，就将发生上述的事情。[3]

因此，不管一件事情简单，比如网球的运动弧线，还是更为复杂，比如一颗鸡蛋的破碎，物理定律都告诉我们，在一个时间方向上发生的事情，至少从理论上来看，是可以反过来发生的。

## 原理和实践

网球运动和鸡蛋的故事告诉我们的不只是自然定律具有时间反演对称性，这些故事还告诉我们，为什么我们在真实的经验世界里看到的许多事情只能朝一个方向发生，反过来则不行。让网球逆着其轨迹运动并不难。拿着它，并以相同大小的速度朝相反方向将其掷出，就这样即可。但使鸡蛋所有的混乱碎片逆着原来的轨迹运动就要困难到不可想象了。我们需要抓住每一片鸡蛋碎片，以相同速度但朝相反的方向同时发送回去。很显然，那远非我们（或者聚齐所有人力物力）所能做到的。

我们找到了我们一直寻求的答案了吗？鸡蛋打碎却无法重新复原（即便两种运动都是物理定律所认可的）的原因是因为其中一种可实现而另一种无法实现吗？答案就是那么简单，就是因为鸡蛋打碎容易——使鸡蛋从桌上滚下去——而使鸡蛋复原难吗？

如果答案是这样的，相信我，我将不会在这里大费周折地讲这么半天了。困难与否确实也是答案的一个重要部分，但整个答案更加奥妙和令人惊奇。在以后的章节中我们将解释这个问题，但这里我们首先需要对这一小节进行更加深入地讨论和了解。为了达到这一目的我们不得不引进熵的概念。

## 熵

在维也纳中央公墓，贝多芬、勃拉姆斯、舒伯特和斯特劳斯的墓穴旁树立着一个刻有s＝klogW方程的墓碑，这一方程就是熵这个强有力的概念的数学公式。这个墓碑的主人就是生活在19世纪、20世纪之交的路德维格·玻尔兹曼，最具洞察力的物理学家之一。1906年，由于糟糕的健康状况和低沉的心情，玻尔兹曼在和妻女在意大利度假时自杀了。具有讽刺意味的是，就在他离世的几个月之后，有实验证实了玻尔兹曼为之毕生热烈维护的思想是正确的。

熵的概念最初是由工业革命时期的科学家们在考虑锅炉和蒸汽机时所提出的，熵的概念促进了热力学领域的发展。通过多年的研究，尤其是在玻尔兹曼的辛勤钻研之后，熵的基本观点被进一步完善起来。玻尔兹曼版本的熵，可用其墓碑上的方程准确的表述，利用统计学原理将构成物理系统的单独组分的数目与系统的整体性质之间联系起来了。[4]

为了感受一下他的思想，想象拆开一本《战争与和平》，将其693页双面纸都高高抛向空中，然后把所有的纸页收集到一堆。[5]当你检查那一堆纸时，你会发现页码混乱的纸张远远比整齐排序的要多。原因是显而易见的。纸张混乱的方式有许多种，而按序排列的方式只有一种。要整齐有序，页码就必须精确排列成1、2、3、4、5、6；一直到1385、1386。其他的排列方式都是无序的。一个简单而又基本的事实是，所有排列方式都是平等的，某件事情发生的方式越多，它发生的可能性就越大。如果某件事情有无数多种的发生方式，就像落地页码的错误排列一样，该事情发生的可能性就有极其大。直观上，我们都可以很好地理解这个问题。如果你买一张彩票，你中奖的方式就只有一种。如果你买了一百万张不同号码的彩票，你就有一百万种中奖的方式，这样你走运的机会就提高了一百万倍。

熵这个概念其实就是该观点的一种具体表述，可以通过数清在物理定律制约下，实现任意给定物理条件的方式的数目来确定相应物理系统的熵的大小。熵越高就意味着实现该物理条件的方式越多，熵越少就意味着方式越少。如果《战争与和平》的页码是按照正确的数字顺序来排列的，则是低熵组合，因为满足标准的只有一种排列方式。如果页码是无序排列，那就是一个高熵组合，很简单的计算就可以告诉我们共有
124552198453778343366002935370498829163361101246389045136887
——约为$10^{1878}$——种不同的无序排列方式。[6]如果你把这些纸张扔向空中，然后再收集成一叠，可以肯定它们将处于无序排列的状态，因

为这种排列方式比唯一的有序排列拥有更高的熵——达到无序排列的方式有很多种。

理论上讲，我们可以运用经典物理学定律来计算将整沓纸扔向空中后每一页所将降落的位置。而且，理论上讲，我们也可以精确预测这些页码的最终排列方式，因而（在量子力学中，情况将有所不同，而那是我们在下一章要讨论的内容）看似没必要依靠诸如此类哪种结果更有可能出现的概率概念。但是统计学确实是强有力且非常有用的工具。如果《战争与和平》只是一本只有几页的小册子的话，我们很快就能成功地完成计算，但是对于真正的《战争与和平》[8]这么做就不可能了。这693张纸随着温和的风飘荡，相互摩擦、滑落、碰撞，最后落到地上，想要追踪这693张纸的精确运动将是一项非常艰辛的工作，远远超出了当今世上最强有力的超级计算机的运算能力。

而且——这点非常关键——即使得出确切的答案也没有什么用处。当你查看这叠纸时，你不会在乎每一页碰巧在哪儿，你感兴趣的是整体效果——它们是否正确排列。如果它们是，那非常好。你可以坐下来像往常一样继续阅读安娜·帕夫洛夫娜和尼古拉·罗斯托夫。但是如果你发现书页的排列乱七八糟，那么你不会在乎这种错误排列具体是怎样的。如果你看到了一种错误的排序方式，你就相当于看到了所有的错误排序方式。除非出于某种古怪的原因，你需要追究每一页的具体下落，否则你甚至都不会注意到是否有人把你那已经混乱的页码搞得更乱。最初的一堆纸就是混乱排列的，即便进一步弄乱也还是混乱的。因此，并不仅仅因为统计学讨论比较容易进行，还因为利用统计学所能得到的结果——混乱或者不混乱——更与我们所真正关心的和需要记下来的事情有关。

这种全局式的思考方式是利用熵来考虑问题的统计学基础的核心。就像任何一张彩票都有中奖的机会一样，《战争与和平》被多次颠倒顺序后，任何一种排列方式都有可能发生。使统计学变得有用武之地的原因在于，我们感兴趣的页码排列方式只是两类：有序和无

序。前一类只有一种（页码正确的排列为1，2，3，4；等等），而后一类则有多种（除正确顺序之外的每一种可能的排列方式）。这两种分类是便于应用的合理分类，因为，就像上文所述，利用这种分类你可以对任何一种页码的排列方式做出全局性的评价。

即便如此，你仍然可能建议对这两种分类进行进一步的区分，比如，只有少数几十页的排列是混乱的，只有第1章的页码排列无序，等等。事实上，考虑这些中间状态的分类有时是很有用的。然而，每一种亚分类中的可能的页码排列方式总数与所有的混乱排列方式总数相比是非常小的。比如说，《战争与和平》第一部分排列混乱的方式总数只不过是所有混乱排列方式总数的百分之一的$10^{-178}$。所以，尽管开始的时候，未装订书所导致的无序页码排列方式可能只属于某种中间状态，而非完全混乱状态，但可以肯定，如果你再三颠倒页码，页码排列顺序最终将展现不出一点规律性。页码排列总是趋向于演变为完全混乱排列的状态，因为这类型的排列方式确实太多了。

《战争与和平》这个例子点出了熵的两个最显著特征。首先，熵是物理系统中无序度的量度。高熵意味着构成系统的组分的许多排列方式毫不起眼，这就相当于说系统处于高度无序状态（当《战争与和平》的页码处于混乱状态时，进一步的颠倒页码顺序几乎不会被大家注意到，因为页码本身就已经处于混乱状态，再颠倒页码也不会产生什么重要影响）。低熵就意味着只有少数一些排列方式显得不起眼，也就相当于说系统处于高度有序状态（当《战争与和平》的页码排列有序时，你很容易就注意到对其顺序所做的任何改动）。第二，由许多组分构成的物理系统（比如说，很多页处于混乱状态的书），有自然演化成更为无序状态的趋向，因为相比于达到有序状态，达到无序状态的方式更多。用熵的语言来说，物理系统倾向于向着高熵状态演化。

当然，要想使熵的概念准确且具有普适性，其物理定义就不能是在使其不变的情况下数清这本或那本书页码的重新排列数目。事实

上，熵的物理定义需要在保持物理系统整体上、大局上不变的情况下，数清其基本组成成分——原子，亚原子粒子，等等——的可能有的排列组合数目。比如在《战争与和平》的例子中，低熵就意味着几乎没有哪次重新排列会不被注意到，因此该系统就处于高度有序状态；而高熵就意味着大量的重排都会显得不起眼，换句话说，整个系统处于无序状态。[28]

让我们来看一个不错的便于说明问题的物理例子，想想先前提到的可口可乐瓶。当气体，比如最初被密封进瓶子里的二氧化碳，传播到房间的每一个角落时，单个分子可能有许多种重排方式，但是这些重排没有什么显著区别。比如说，当你挥动胳膊时，二氧化碳分子将会来回穿梭，迅速改变位置和速度。但从整体上看，分子的调整不会带来整体性质上的变化。在你挥动胳膊之前分子是均匀分布的，挥动胳膊之后仍然是均匀分布的。气体的这种均匀分布状态对于分子的大量重排方式是不敏感的，这正是所谓的高熵状态。相比而言，如果气体分布在较小的空间内，比如说瓶子内，或被障碍物密封在房间墙角，就会出现有意义的低熵状态。原因很简单。正如一本薄薄的书的页码只可能有几种排列方式，小地方只能为分子的排列提供一点点空间，因而也就只会产生很少的排列方式。

但当你拧开瓶盖或是挪开障碍物时，你就为气体分子打开了一个全新的世界，它们开始运动、碰撞，很快播散到房间的各个角落。为什么呢？这和《战争与和平》问题中的统计学推演是一样的。毫无疑问，一些分子经过碰撞将会离开最初的气体团或使一些刚刚离开的气体分子又被撞回来。但因为房间的体积远远超过了最初的气体团，如果它们分散开来，分子将会有更多种的排列方式。因此，气体从最初的低熵状态——气体聚集在一个小区域内，自然演化到高熵状态——气体均匀地分布在更大的空间内。一旦气体达到这种均匀状态，将一直维持高熵状态：运动和碰撞会使分子四处移动，从而造成一种又一种的重排方式，但大部分重排方式都不会影响气体的全局性的、整体

性质。而这就意味此时处于高熵状态。[9]

理论上，就像《战争与和平》的页码一样，我们能用经典物理学定律来精确确定在某一特定时刻每一个二氧化碳分子的位置。但由于二氧化碳的分子数目太大了——一个可乐瓶里大约有$10^{24}$个——进行这样的计算事实上是根本不可能的。但不管通过什么方式，即使我们做到了，手里拿着一张记有亿亿亿个粒子的速度和位置的单子，对于我们了解分子是如何分布的并没有多大的意义。把焦点集中在全局性的统计特点上——气体四散分布或集中在一起，也就是说，气体处于高熵状态还是低熵状态——才更富有启发性。

熵——第二定律和时间之箭

物理系统趋向于高熵状态就是所谓的热力学第二定律（第一定律就是熟悉的能量守恒定律）。如上文所说，该定律的基础是简单的统计学推演：系统有更多的方式达到高熵状态，“更多的方式”就意味着系统更有可能演化为某种高熵状态。注意，尽管从传统的意义上看，这并不是一条定律；这是因为，尽管极为罕见并且几乎不可能发生，但是诸如某物从高熵状态演化到低熵状态的事件是有可能会发生的。当你把一堆混乱的纸扔向空中，然后收集成一小摞时，它们有可能按完美的页码顺序排放。你大概不会想在这种结果上下大赌注，但它的确是可能发生的。运动和碰撞也有可能碰巧使所有分散的二氧化碳分子一起移动，嗖的一下全都返回到了打开的可乐瓶中。你当然不会凝神静气睁大双眼等待着这种结果的发生，但它的确是可能发生的。[10]

《战争与和平》庞大的页数以及房间里气体分子的巨大数目使得有序和无序排列之间的熵的差别如此之大，而这使得低熵结果很难发生。如果你把两张双面纸一次次扔向空中，你将会发现它们落地时按正确顺序排列的次数为所扔次数的12.5%。3页纸的话这个概率将减小为2%，4页纸将是0.3%，5页纸将是0.03%，6页纸将是0.002%，10页纸将是0.000000027%，693页纸扔向空中而落回地面时正确排

列的概率就更小了——小数点后包含了许许多多的零——我确信出版商不想浪费一页纸来把它详尽地列出来。类似的，如果你只把两个气体分子肩并肩地放进空可乐瓶里，你将会发现在室温下，平均每隔几秒钟，随机运动就会把它们弄到一起（相距1毫米之内）一次。但如果是3个分子，你就不得不等好几天，如果是4个分子就得好几年，如果是最初气体团里有亿亿亿个分子，那就不得不花比现在宇宙年龄还长的时间来等待随机运动使它们同时聚集到一个小而有序的气体团中。比死亡和纳税还要可靠的是，我们可以相信，一个具有很多组分的系统倾向于向无序状态演化。

虽然不会马上看清，但是我得说我们现在遇到了一个有趣的问题。热力学第二定律似乎为我们带来了时间之箭，这根时间之箭只有当物理系统拥有相当多的组分时才会出现。如果你看到一部片子正在放映两个二氧化碳分子被放置在一个小盒子里（示踪计显示了每个分子的运动轨迹），你将很难辨别片子到底是在正着放还是反着放。两个分子飞来飞去，一会儿一起运动，一会儿分开运动，但它们不会展现出任何整体的迹象，可以使我们辨别出时间的方向。然而，如果你看到一部片子正在放映$10^{24}$个分子聚集在盒子里（就像一团小的高密度分子云），你很容易就能辨别出片子是正着放还是反着放：几乎不用怀疑，时间前进的方向就是气体分子变得越来越均匀，从而达到越来越高的熵的方向。相反，如果电影正在播放均匀分布的分子嗖的一声集中到一小团的场景，你立刻就会意识到片子放反了。

这种推演适用于我们在日常生活中遇到的所有事情——即有很多成分组成的事物：时间之箭的箭头指向熵增长的方向。如果你在片子中看到吧台上有一杯冰水混合物，你就可以通过查看冰是否在融化来判断时间之箭的方向——水分子扩散到整个杯中，因此达到了更高熵的状态。如果你在片子中看到一个破碎的鸡蛋，通过检查鸡蛋的成分是否越来越处于无序状态——鸡蛋破碎出来就是向着高熵状态——来确定时间的方向是否向前。

正如你所见，熵的概念为我们先前发现的“难易”结论提供了一个精确的版本。《战争与和平》页码容易弄乱是因为有如此多种无序排列方式。这些页码很难按恰好的顺序降落，因为这需要上百张纸降落时恰好按照托尔斯泰的意愿。一个鸡蛋很容易破碎因为有如此多的破碎方式。一个破碎的鸡蛋很难会记起来，因为无数个破碎的鸡蛋成分必须以和谐的步调移动才能形成放在桌上的一个独立完整的鸡蛋。对于由多种成分构成的物质而言，从低熵状态达到高熵状态——从有序到无序——是容易的，因此它总在发生。从高熵状态到低熵状态——从无序到有序——是非常难的，因此很少发生。

请注意，这里所说的熵的方向并不是完全严格的；而且也没有声明时间方向的定义就是100%正确的。相反，也有不少方法可以允许这样或那样的过程反向发生。因为热力学第二定律声明熵的增长只是一种统计学的可能性，而不是大自然中不可避免的事实，它允许存在一点这样的可能性：页码落下时能恰好按顺序排列，气体分子能聚合起来并重新回到瓶子里，碎鸡蛋能汇聚起来。通过熵的数学公式，热力学第二定律精确说明了这些事件在统计学上的不可能性是多大（注意，前一小节中那长达一页的巨大数字反映了这些书页无序降落的可能性有多大），但同时这也意味着它们可能发生，只是概率非常小而已。

看起来这个故事很有说服力。统计学和概率论证为我们带来了热力学第二定律。接着，第二定律为我们所谓的过去和未来提供了直观上的区别。熵也为我们日常生活中的现象提供了一种实用的解释，那些由大量组分构成的事物，以这种方式开头而以那种方式结尾，而我们从未看到它们以那种方式开头而以这种方式结尾。经过许多年的努力——也多亏了像开尔文勋爵、约瑟夫·洛施密特、亨利·庞加莱、S.H.勃柏利，欧内斯特·切梅罗以及威拉德·吉布斯等物理学家的重要贡献——路德维格·玻尔兹曼开始意识到，有关时间之箭的整个故事更加令人惊奇。玻尔兹曼意识到，虽然熵阐明了这个谜团的重要方面，但并

没有回答为什么过去和未来看起来如此不同。正相反，熵以一种重要的方式精炼了这一问题，而这为我们带来了一个出乎意料的结论。

熵——过去和未来

在前文中，通过将我们日常生活中的事实与经典物理中牛顿定律的性质相比较，我们提出了有关过去和未来的难题。我们发现，我们每天所不断经历的事情在时间上具有很明显的方向性，但物理定律却平等的对待时间上的所谓将来和过去。由于物理定律没有表明时间具有方向性，也没有明确声明“只能沿着时间方向上运用这些定律，不可逆向使用”，于是我们不得不追问：如果以经验为基础的定律认为时间在方向上是对称的，为什么这些经验本身却具有时间上的倾向性，总是在一个方向上发生而不会在其他方向上发生呢？我们所观测到和体验到的时间的方向性来自哪里呢？

在上一小节中，看似我们已经通过热力学第二定律取得了一定进展，该定律清楚地将未来定为熵增多的方向。但进一步思考后，我们发现事情并不是这么简单。值得注意的是，我们在关于熵和第二定律的讨论中，并未以任何方式修改经典物理学定律。相反，我们所做的一切，只是在“全局性的”统计框架中运用这些定律：我们忽略了微妙的细节（《战争与和平》未装订页码的准确顺序，鸡蛋组分的精确位置和速度，可乐瓶中二氧化碳分子的精确位置和速度），而把焦点集中在全局性的整体特点上（页码的有序排列和无序排列，鸡蛋的破碎和汇集，气体分子的广布和聚集）。我们发现当物理系统足够复杂（由许多页码组成的书，会破碎成很多片的易碎物品，由许多分子组成的气体）时，其组分处于有序还是无序状态，熵的区别是很大的。这也就意味着，系统很有可能会从低熵状态演变到高熵状态，这正是热力学第二定律的粗略描述。但需要注意的是，第二定律是派生出来的：它只是将概率推演应用于牛顿运动定律时得到的结果。

这就导致一个简单而又令人惊奇的问题：既然牛顿定律没有内在

的时间方向，我们用以论证物理系统会沿着未来的方向从低熵向高熵状态演化的全部推演，也同样适用于过去。又一次，由于深层次的基本物理定律具有时间反演对称性，因而它们无法区分所谓的过去和未来。就像在漆黑的外太空中没有标牌指示哪个方向是上，哪个方向是下一样，经典物理学中没有任何定律说明时间上哪个方向是未来，哪个方向是过去。定律并未提供时间方向；它们对时间方向上的区别完全不敏感。因为运动定律着眼点在于事物的改变——既可以朝向所谓的未来，也可以朝向所谓的过去——热力学第二定律背后的统计或概率推演同时适用于两个时间方向。因此，一个物理系统的熵，不仅存在很大的概率在所谓的未来会变高，也有很大的概率在所谓的过去曾非常高。如图6.2所示。

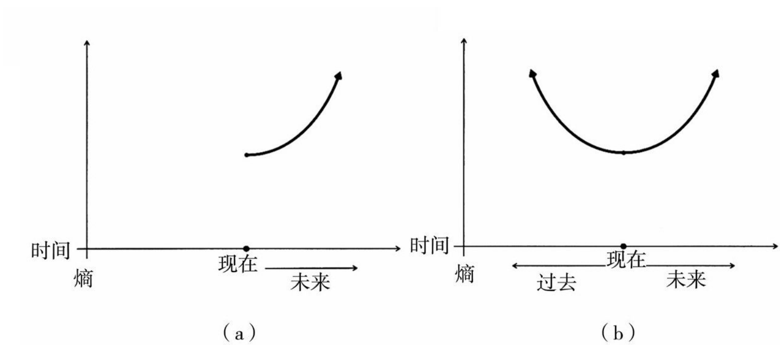



图6.2　（a）：如通常所言，热力学第二定律告诉我们熵在未来会随着时间流逝而增加。（b）：已知的自然定律并未对未来和过去区别对待，因而，热力学第二定律实际上告诉我们，熵在向着未来和过去这两个方向上都是增加的。

这一点对下面的讨论非常关键，但也是富有欺骗性的微妙之处。通常的误解是，根据热力学第二定律，如果熵会在朝向未来的方向增加，那么熵当然就会在朝过去的方向上降低。但这正是微妙之处。第二定律实际上是说，在任一给定时刻，如果物理系统碰巧没有拥有最大的熵，那它很可能在下一刻会拥有且在前一刻曾拥有更高的熵。这正是图6.2（b）的内容。由于定律并不区分过去和未来，时间上对称性是不可避免的。

这是重要的一课，它告诉我们熵所带来的时间之箭是双向的。从任一明确时刻起，熵增的箭头既会朝向未来也会朝向过去，这就使得很难把熵作为经验时间具有单向性的解释了。

想象一下熵的双向性在具体例子中的含义。比如在暖和的某一天，你看到一杯水中有一块部分融化的冰块，那么你就可以确信半小时之后冰块会融化得更厉害，因为只有它们融化得更厉害，熵才会变得更高。[11]但是，你也应同样确信，这块冰半小时之前融化得更厉害，因为完全一样的统计推演告诉我们熵会朝着过去的方向增加。同样的结论也适用于每天我们遇到的无数例子。既然你确信熵会朝着未来的方向增加——四散的气体分子在未来会继续扩散，页码部分混乱的书页会变得更加混乱——那你也应当同样相信熵在过去应该更高。

问题在于，这些结论中的一半，看起来明显是错误的。当熵的推演用于一个时间方向，朝向我们所谓的未来时，就会产生准确而合理的结论；但当应用于我们所谓的过去方向时，就会明显产生不准确而且看起来非常荒谬的结论。杯中带有冰块的水通常不可能一开始就是一杯完全没有冰块的水，然后水分子聚集起来冻成冰块，然后再一次融化。《战争与和平》未装订的书页通常不会开始于无序排列，继而被扔向空中后就变得没有以前混乱了，它只能越来越混乱。再返回厨房，鸡蛋通常不会一开始就是破碎的，然后再集合起来形成一个完整的鸡蛋，它只能由完整的鸡蛋打碎。

难道，它们竟可以这样吗？

跟着数学走

几个世纪的科学研究表明，数学为分析宇宙提供了有力而敏锐的语言。确实，现代科学的历史中，满是数学做出貌似与直觉和经验相违背预测（比如宇宙存在黑洞以及反物质，间隔很远的粒子可以发生纠缠，等等），但实验与观测却最终证实数学预言的例子。这样的发

展历程在理论物理文化中留下了深深的烙印。物理学家们开始意识到，数学，如果使用得足够小心，将是通向真理的可靠路径。

因此，当自然定律的数学分析表明，某一时刻的熵既会朝着未来增加，也会朝着过去增加时，物理学家们并不会立即驳回它。相反，一些类似于物理学家的希波克拉底誓言的信条激励着研究者们对人类体验的明显事实保持深刻而健康的怀疑态度，带着这样的怀疑态度，孜孜不倦地跟着数学走，看看它将把我们带到哪里。只有那时，我们才能正确评价和诠释物理定律和常识之间的不匹配之处。

为了达到这个目标，想象一下现在是晚上10：30，半小时以来你一直盯着一杯冰水（这是酒吧里一个悠闲的夜晚）观测冰块慢慢融化成小块，最后乃至不见。你毫不怀疑半小时之前男服务员往你杯子里放了几个完整的冰块；你毫不怀疑是因为你相信你的记忆力。即使偶尔你对于半小时之内所发生事情的信心动摇了，你可以问问过道里的小伙，他们看见了冰块的融化（这确实是酒吧里一个悠闲的夜晚），或者看看酒吧监视器摄的录像，它们都可以使你相信你的记忆没有问题。如果你问自己，接下来的半小时内，冰块将会怎样，你可能会想到它们将继续融化。如果你非常熟悉熵的概念，你将把你的预测解释为从你看到冰块的那一刻起，那时刚好是10：30，向着未来的方向，熵将不断增长。所有这些都很合理，并且与你的直觉和经验相符。

但正像我们所看到的，有关熵的这样的推演——简单地认为事物更有可能达到无序的状态是因为有更多种方式可以达到无序状态，这种推演在解释事物是如何向未来发展时无疑是强有力的——告诉我们，熵在过去也有可能更高。这就意味着你在晚上10：30看到的部分融化的冰实际上在早些时候融化得更加厉害；这也就是说，在晚上10：00时，它们还不是固体冰块，而是从那时到晚上10：30这段时间，它们在室温的水中慢慢地集合起来形成冰块；正如10：30到11：00这段时间它们会慢慢融化成室温的水一样。

毫无疑问，这听起来非常古怪——或者你会说这太荒谬了。如果是真的，不仅需要杯子里室温下的水分子会自发集合起来形成部分融化的冰块，而且监视器上的数码和你大脑中的神经元，以及过道里小伙子的神经元，都需要在晚上10：30以前有所调整，以便证明水曾形成完整的冰块，即便它实际上从不存在。但是，这种古怪的结论却是在物理定律所展现的时间对称性背景下，应用值得信赖的有关熵的思考——你曾毫不犹豫地用这种思想解释你在晚上10：30看到的部分融化的冰到11：00的这段时间里继续融化——而得出的结论。这就是基本运动定律不能区分过去和未来而造成的麻烦，这些定律的数学以完全相同的方式处理某一给定时刻的过去和未来。[12]

放心好了，我们很快就能找到方向，逃出由于运用熵来思考问题时平等地看待过去和未来而陷入的窘境；我不会试图让你相信存在于你的记忆和记录中的过去从未真的发生过（对《黑客帝国》迷们我只能说抱歉了）。但是，我们将会发现，弄清直觉和数学定律之间的不同之处是极为有益的。因此，我们继续抓着这条线索。

一片沼泽地

直觉让你觉得高熵的过去不够满意，因为当用通常的事件向前发展的方式来看时，高熵的过去意味着有序度会自发增加：水分子自发的冷到0摄氏度然后变成冰，大脑自发的获得不曾发生过的事情的记忆，录像机自发的产生从未发生过的事情的图像，等等，所有的这些都极不可能发生——一种连奥利弗·斯通29都会嘲笑的解释过去的方式。在这一点上，物理定律和熵的数学公式与你的直觉完全一致。事件的这种发生顺序，当按晚上10：00到10：30的时间方向来看时，与热力学第二定律相违背——熵会减少——因此，虽然不是不可能，但却极其不可能发生。

通过对比，你的直觉和经验告诉你，更有可能的事件顺序是，晚上10：00，冰块很完整，到了现在，晚上10：30，你盯着的玻璃杯

中的冰块部分融化了。但在这一点上，物理定律和熵的数学公式只与你的期望部分相符。数学公式和你的直觉相一致的是，如果晚上10：00时真的存在完整的冰块，则最有可能的事件顺序是，到了晚上10：30，你一直盯着的杯中的冰部分融化了：这一熵增的结果既与热力学第二定律又与你的经验相符。但数学和直觉有所区别之处在于，我们的直觉，不像数学，没法考虑也不会考虑这样的可能性，即在假定你在晚上10：30时看到冰块部分融化——被我们视为无可辩驳的、可靠的事实——的确发生了的情况下，晚上10：00时真的存在完整的冰块。

这一点非常重要，我们来解释一下。热力学第二定律的核心内容在于，物理学系统强烈的倾向于处于高熵状态，因为这种状态可以通过多种方式实现。并且一旦物体处于高熵状态时，就有很大的倾向性继续保持在该状态。高熵是自然形成的状态。你不需要惊讶或感到有必要解释为什么物理系统会处于高熵状态。这样的状态是正常的。相反，需要解释的是为什么给定的物理系统处于有序状态，一种低熵状态，这种状态是不正常的。它们当然会发生。但从熵的角度来看，这种有序状态属于违背常规的少数情况，需要有所解释。因此，上一节中我们毫不怀疑就相信的一个事实——你会在晚上10：30时看到处于低熵状态部分融化的冰块——实际是需要解释的。

从概率的角度看，借助更低熵的状态来解释低熵状态是十分荒唐的；更低熵的状态指的是，晚上10：00时在更为原始、有序的环境里你所观测到的更加有序、更为完整的冰块。与之不同的是，事情更有可能开始于毫无奇特之处的、十分平常的、高熵的状态：一杯完全没有冰块的纯净水。然后，通过一种可能性不高但偶尔会发生的统计涨落，这杯水背离了热力学第二定律，演化为含有部分融化冰块的相对低熵状态。这种演化，尽管需要少见且不熟悉的物理过程，但却完全规避了更低熵、更不可能发生、更为少见的、拥有完整冰块的状态。在晚上10：00到10：30这段时间的每一时刻，这种听起来有点奇怪

的演化过程比正常的冰融化过程拥有更高的熵，就像你在图6.3中所看到的，这样，它就以一种比完整冰块融化的可能性更大的方式——更有可能发生——实现了晚上10：30时所观测到的现象。[13]这才是关键之所在。30

对于玻尔兹曼而言，意识到整个宇宙可归结为同样的分析只是迈进了一小步。当你环顾宇宙时，你所看到的会是大量的生物组织、化学结构和物理序列。虽然整个宇宙可以处于一种完全无组织的混乱状态，但它不是这样。这是为什么呢？这种有序来自哪里呢？就像冰块一样，从概率的立场看，我们今天所看到的宇宙不可能从遥远的过去更加有序的状态（这种可能性就更小了）慢慢地演化成今天的样子。实际上，由于宇宙的组成部分如此之多，有序和无序状态的规模就被放大了。因此酒吧里的真实状况就是整个宇宙状况的真实写照：更有可能的是——毫无疑问极有可能——我们今天所见的整个宇宙来自于一种正常的、毫不出奇的、高熵的、完全混乱的状态的统计学涨落。

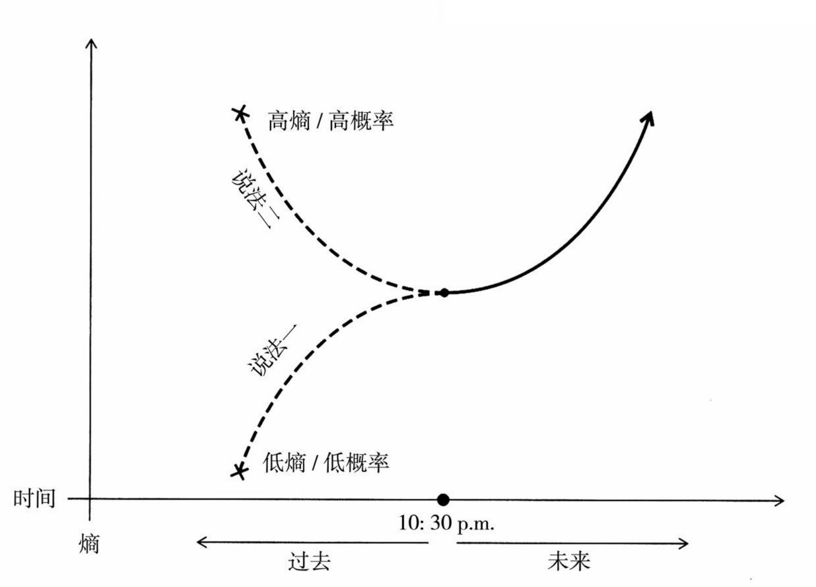



图6.3　关于冰块如何变成现在（晚上10：30）这种部分消融状态有两种说法，这里对比一下

这两种说法。说法一与你记忆中的冰融化过程一致，但是需要冰块在晚上10：00开始融化的时候处于熵相对低些的状态；说法二则要挑战你的记忆了，晚上10：30你看到的冰处于部分融化状态，但这种状态却开始于晚上10：00时一个高熵、高度无序的状态。在向着晚上10：30时的状态演化的过程中，说法二的可能性更大一些——因为，如你在图中看到的那样，其熵更高一些——说法二更为符合统计要求。

尝试着用这种方式来思考一下整个问题：如果你将一把硬币一次又一次的抛向空中，它们迟早都会正面落地。如果你有足够多的耐心一次又一次地把《战争与和平》的混乱的页码扔向空中，它们迟早都会以正确的顺序落地。如果你拿一瓶跑了气的可乐等待，随机运动的二氧化碳分子迟早都会重新回到瓶子里的。对于玻尔兹曼的批判者而言，如果宇宙等待足够长的时间——几乎是永恒的等待——其普通的、高熵的、高概率的、完全混乱的状态将通过粒子的移动、碰撞、随机运动和辐射，最终碰巧融合形成我们现在所看到的结构。我们的身体和大脑——储存着记忆、知识和技能——完全形成于混沌，甚至记忆中的过去也可能从未真的发生过。我们所了解的每一件事物，我们所看重的每一件东西，不过是稀有但却意料之内会偶尔发生的统计涨落，这种涨落会暂时打破近似永恒的无序状态。如图6.4所示。

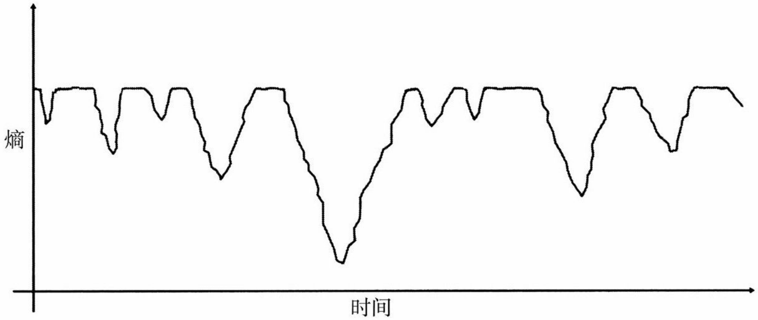



图6.4　宇宙总的熵随时间变化的示意图。从图中我们可以看到，宇宙在大多数时间里都处于完全无序的状态——高熵态——状态上的起伏很频繁，有序度变化很大，也会频繁地变到低熵态。熵上变化越大的涨落越不可能发生。熵上显著的变化——到今天宇宙这种有序度的变化——极其不可能，发生的次数很少。

## 回头看看

许多年前当我第一次了解到这种想法时，我很是震惊。在那之前，我曾经认为我对熵的概念理解得还不错，但事实是，按着我学习的教科书的思路，我只能想到将熵应用于未来。而且，正如我们刚刚看到的那样，若我们将熵的概念用于关于未来的讨论，则一切都和我们的直觉和经验相符，而一旦将熵的概念用于讨论过去，则一切又与我们的直觉和经验相矛盾。这种感觉或许还没差到突然得知自己被相交多年的老朋友出卖了那么糟糕，但是对我来说，其实也差不多。

然而，有时候我们的结论最好不要下得太早，熵的表现未如预期恰恰为我们带来了一个很重要的例子。你或许正在想，我们所熟悉的各种思想突然变得面目全非，这种事情一时真难消化。而且，关于宇宙的这种解释并不“仅仅”动摇了那些我们认为真实又重要的东西，它还留下了一些尚未有答案的重要问题。比如说，今天宇宙有序度越高——图6.4中的凹陷越深——使其发生的统计涨落就越让人觉得不可思议且不可能。因此，如果宇宙有什么捷径可走，在不需要实际上的那么高有序度的情况下，使事物多少看起来像是我们现在所看到的这样，那么概率上的原因就会使我们相信它真的会那么做。但当我们研究宇宙时，发现错失的机会实在太多了，因为很多事物的有序度都比其本来需要的多。如果麦克尔·杰克逊从没有灌制过《战栗》这张唱片，这张唱片分布在世界各地的数百万份拷贝的存在只不过是朝向低熵的反常涨落，那么相对来说，拷贝只有100万份或50万份甚至只有几份的话，这种反常涨落就显得没那么严重。如果进化从未发生，人类的存在只不过是朝向低熵的反常涨落，那么，根本就不存在证明进化的化石的话就会使涨落没那么严重。如果大爆炸从未发生，我们所看到的数千亿之多的星系只不过是朝向低熵的反常涨落，那么，星系的数目只有500亿，5000，或是更少，甚至只有一个的话就会使反常涨落没那么严重。所以，如果有人认为我们的宇宙只不过是统计学涨落这样的想法——一次幸运的偶然事件——正确的话，那他就需要解释清楚宇宙怎样以及为什么会走得如此之远，以至于达到了今天这种极低熵的状态。

更进一步，如果你真的不能相信记忆和记录，那么你也没办法相信物理定律。它们的正确性取决于数不清的大量实验，而这些实验的结果却又需要记忆和记录的证明。因此，所有基于公认物理定律的时间反演对称性的思考都会有问题，从而干扰我们对熵的理解，破坏当前讨论的整个基础。如果我们相信我们所认识的宇宙只不过是完全无序状态的罕有但偶尔也会发生的统计涨落，那么，我们很快就会陷入困境，我们会发现我们将丧失所有的思维结果，包括一开始为我们带来这种古怪解释的一系列思考。[31]

因此，将怀疑放在一边，努力跟着物理定律和熵的数学公式走——这些概念结合起来会告诉我们，从任意给定时刻开始，无序度很有可能既会朝着未来也会朝着过去的方向增长——我们很快就会掉入陷阱。虽然听起来不怎么美妙，但这件事的确不错，原因有两点。第一，它准确地说明了为什么怀疑记忆和记录——直觉上我们会鄙视的东西——不合理。第二，当我们发现整个分析框架处于崩溃的边缘时，我们被迫认识到，在我们的推理过程中，某些重要的东西必定被漏掉了。

因此，为了避开思维上的深渊，我们问自己：除了熵和自然定律的时间对称性外，我们还需要有哪些思想或概念，才能使我们重新相信自己的记忆和记录——室温下的冰块会融化而不是不融化，奶油和咖啡会混到一起而不会自然分开，鸡蛋会破碎而不会重新组合起来？简而言之，如果我们用熵在未来方向不断增长而在过去方向降低的说法来解释时空中事件发展的不对称性，我们将会得到什么样的结果呢？有这种可能性吗？

有。但除非初始时事物非常特殊。[14]

鸡蛋、鸡和大爆炸

为了弄清楚这是什么意思，我们来看看前面提到的，低熵的、完

整的鸡蛋。这种低熵的物理系统是如何形成的呢？如果我们能信任记忆和记录的话，我们就知道答案了。鸡蛋来源于一只鸡。鸡来源于鸡蛋，而鸡蛋又来源于鸡，鸡又来源于鸡蛋，如此反复。但是，正如英国数学家罗杰·彭罗斯特别强调的那样，[15]这个鸡和鸡蛋的故事实际上教会了我们一些更为深刻的东西并使一些问题更为明确。

鸡，或其他生物，是一种令人惊讶的高度有序的物理系统。这种组织性来自哪里并且又是如何维持的呢？鸡仍然存在，并且可以靠不断生蛋、吃食以及呼吸继续存在下去。食物和氧气为生物提取所需的能量提供了原材料。如果我们要真正理解究竟是怎么回事的话，这种能量的一个重要特点就不得不强调一下。在鸡的一生当中，鸡通过摄取食物获得能量，然后又将能量以新陈代谢和日常活动所产生的热量和废物的形式排放到周围的环境中。如果没有这种能量摄取和释放的平衡，鸡将越来越笨重。

问题的关键在于，各种形式的能量并不一样。鸡以热量释放到环境中的能量是高度无序的——这些热量常常导致周围的空气分子的震动碰撞变得比先前剧烈。这种能量的熵很高——这些能量不断散发，并与环境混合在一起——因此不能轻易利用。相反，鸡从食物中摄取的能量的熵则很低，因而很容易用于重要的维持生命的活动。因此鸡，事实上也包括每一种形式的生命，都在摄取低熵能量释放高熵能量。

认识到这一点又会发现另一个问题。鸡蛋的低熵源自哪里。鸡的能源食物又是如何拥有如此低的熵的？我们应如何解释这种反常的有序？如果食物是动物来源的话，我们又回到了最初的问题：动物是如何拥有低熵的？但如果我们追踪食物链，我们最终将发现动物（比如我）只吃植物。植物和果蔬产品又是如何维持低熵的？在光的作用下，植物通过光合作用将周围空气中的二氧化碳转化成氧气和碳水化合物，氧气释放到空气中，而碳水化合物被植物吸收利用以生长繁殖。因此我们能将低熵的、非动物性的能源追踪到太阳那里。

这又进一步引起了解释低熵的另一问题：高度有序的太阳来自哪里？太阳形成于50亿年前，它最初是由弥漫的气体团在其组成成分相互之间的引力作用下不断地旋转、聚集而形成的。当气体团密度变大时，一个部分施加于另一个部分的引力就会增强，从而造成气体团进一步向自身塌陷。当引力将气体团挤压得越来越紧时，气体团就会变得越来越热。最终，气体团的温度如此之高以至于引发了核反应，从而不断向外辐射热量以阻止引力对气体团的引力压缩作用。这样，一个高温、稳定、明亮燃烧着的恒星就诞生了。

那么，分散的气体团又来自哪里呢？它可能来源于较老恒星的残余物，当恒星的生命走向尽头时，会爆发变成超新星，并将其物质喷向太空。那么，形成早期恒星的分散气体又来自哪里呢？我们相信这些气体是在大爆炸之后形成的。我们有关宇宙起源的最精确理论——我们最为精妙的宇宙学理论——告诉我们，当宇宙的年龄只有几分钟时，宇宙间充满了由约是75%的氢，23%的氦，少量的氘和锂组成的近乎均匀的高温气体。最关键的一点是，充满宇宙的这些气体的熵是非常低的。诞生于大爆炸的宇宙始于低熵状态，这种状态正是我们现在看到的有序态的起源。换句话说，现在的有序态是宇宙的遗迹。让我们更为详尽地讨论一下这一重要的思想吧。

熵与引力

理论和观测都表明在大爆炸后的几分钟内，原初气体均匀地分布在年轻的宇宙中，你可能会想，考虑到先前讨论过的可乐和二氧化碳分子，原始气体会处于高熵的无序状态。但事实并非如此。早前我们讨论熵的时候完全忽略了引力的影响，当时这样做是十分明智的，因为当少量的气体从可乐瓶里跑出来时引力几乎不起什么作用。在这一假设下，我们发现均匀分布的气体会有很高的熵。但当引力起作用时，情况就不一样了。引力是一种无所不在的吸引力；因此，如果有很大质量的气体，那么每一部分的气体对其他部分的气体有吸引力，而这会使得气体聚集成团，就像蜡纸上的表面张力会使其上的水凝结

成小水滴。当引力起作用时，在早期宇宙的高密度状态下引力就会起作用，团状结构——而不是均匀分布——才是常态；气体会倾向朝这种状态演化，如图6.5所示。

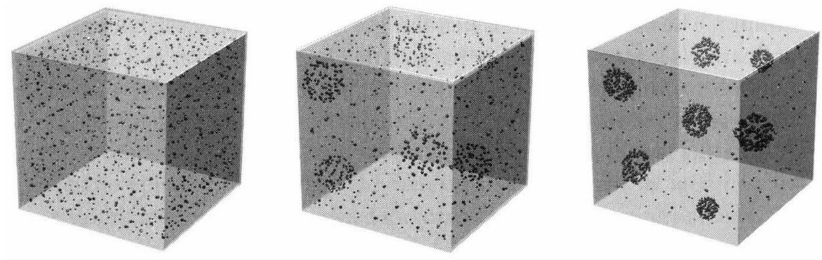


图6.5　对于巨大体积的气体来说，当引力起作用时，原子和分子会从一个平滑均匀的分布演化到具有较大较密团状结构的分布。

虽然气体成团比最初的四散状态更为有序——就像玩具整齐地放在游戏室的箱子里，总比玩具撒得到处都是更为有序——但在计算熵的时候你还是需要将所有源头的贡献都考虑进去。在游戏室的例子中，将被扔得四处都是的玩具堆放到箱子和抽屉里，会使熵减少；而家长花了几个小时收拾房间、整理玩具又会消耗脂肪产生热量，这个过程又会造成熵增；不过，后者的熵增足以补偿前者的熵减。类似的，对于最初四散的气体而言，你会发现气体在有序聚集的过程中熵会减少，而气体在压缩过程中所产生的热量以及核反应过程发生时释放的大量热量和光会导致熵的增加，这里的熵增也同样大过熵减。

这一点非常重要，但时常会被人们忽略。朝无序状态的演化虽然不可抗拒，但这并非意味着像恒星和行星那样的有序结构，或者像植物和动物那样的有序生命形式，不能在这个过程中形成。它们可以形成，而且显而易见的确就是这样。热力学第二定律带来的结果是，在形成有序结构的过程中会生成更多的无序。即使某些成分变得更加有序，熵的账本上仍在不断赢利。在自然界的基本力中，引力对熵的这个特点利用得最为充分。因为引力不仅在长距离上起作用，还无所不在，它引发了有序团块结构——恒星——的形成，而恒星又会发出我

们在晴朗的夜空中可以看到的光，所有这一切的净效果就是造成了熵的增加。

气体团压缩得越厉害、密度越大、质量越重，其整体的熵就越大。黑洞——在引力的团聚和压缩作用下宇宙中所能有的最极端形态——将这一点发挥到极致。黑洞的引力如此之强，以至于没有任何东西，即便是光，可以从中逃逸，这就是黑洞之所以黑的原因。因此，不同于普通的恒星，黑洞顽守着其所产生的所有熵：没有任何东西能逃脱黑洞强大引力的吸引。[16]事实上，正如我们将在第16章中所讨论的那样，宇宙中没有任何东西能比黑洞[32]包含的无序还多，也就是说没有任何东西能有比黑洞更多的熵。这倒与我们的直觉相符：高熵意味着在物体形态不发生改变的情况下其组成成分的重排数目更多。既然我们不可能看见黑洞内部，我们也就不可能探测到其组分——不管那些成分是什么——的任何重排，所以黑洞的熵必然最大。当引力将其肌肉收缩到极限时，它就成了已知宇宙中最有效的熵生成器。

现在，我们的猎物终于要停下来了。有序和低熵的终极起源一定是大爆炸本身。在宇宙的最初时刻，还没有像黑洞这样超大的熵容器存在，我们只能从概率的角度考虑，由于某些原因，新生的宇宙充满了热而均匀的氢气和氦气混合物。尽管这种结构本身熵很高，但由于密度很低，所以我们可以忽略引力，而引力不能被忽略时情况就全然不同了；因此，这种均匀气体的熵非常低。与黑洞相比，这些分散而近乎均匀的气体处于非同于寻常的低熵状态。从那时起，根据热力学第二定律，宇宙的总熵渐渐变得越来越高；总的净无序度也在渐渐增长。大约过了10亿年后，在引力的作用下，原初气体不断聚集，最终形成了恒星、星系，其中较轻的形成了行星。于是，至少有一颗这样的行星，它的附近有一颗恒星，这颗恒星提供了相对低熵的能源，这些低熵的能源使得低熵的生命形式得以演化，在这些低熵的生命形式中最终有一只鸡下了一个蛋，而这只蛋几经周折现在摆放在你厨房的餐桌上，令你气愤的是鸡蛋继续进行着向高熵状态演化，它从桌上掉

下来，在地上摔碎了。鸡蛋之所以摔碎而不是聚集起来，是因为它在朝着高熵状态前进，而高熵状态是由宇宙诞生时的低熵状态引起的。宇宙诞生时令人难以置信的有序态正是一切的开始，从那时起我们一直都生活在这种渐渐向高熵状态演变的宇宙中。

这就是串联起整个这一章的神奇线索。摔碎的鸡蛋告诉了我们一些有关大爆炸的深刻东西。它告诉我们大爆炸带来了一个高度有序的新生宇宙。

同样的思想也可用于许多其他例子。把一本未装订的《战争与和平》扔向空中会导致高熵状态，是因为这本书开始于一种高度有序的低熵形态。其初始的有序形态为熵的增加做好了准备。相反，如果一开始这些页码并没有按顺序排好，则将其扔向空中时，熵不会发生多大变化。又一次，我们不得不提出这个问题：这些书页是怎样变得如此有序的呢？托尔斯泰按一定的顺序写作，印刷工和装订工按照他的原意进行印刷装订。托尔斯泰和这本书的生产者那高度有序的身体和意识允许他们创造出这样一本高度有序的书，而其身体和意识的高度有序则可以用我们解释鸡蛋时的思维来解释，这就又一次使我们回到了大爆炸。你在晚上10：30看到的部分融化的冰块又怎样呢？现在我们姑且相信记忆和记录，你印象中晚上10：00时服务员曾把完整的冰块放进了你的杯子里。他从冰箱里取出了冰块，冰箱是由聪明的工程师设计，天才的机械师制造出来的，他们之所以能创造出如此高度有序的东西是因为他们本身就是高度有序的生命。又一次，我们发现无序态可以追溯到高度有序的宇宙起源。

关键输入

我们所能得到的启示是，我们可以相信记忆中的过去处于低熵而不是高熵状态，只要大爆炸——创造宇宙的过程或事件——所创造的宇宙一开始处于极不寻常的低熵高度有序状态。如果没有关键输入，我们较早前的认识——在任意给定时刻，熵都会既朝未来的方向又朝

过去的方向增长——将使我们得出这样一个结论，即我们所见的所有有序态都源于普通的高熵无序态的偶然涨落，我们已经看到，这样一个结论恰恰破坏了推出该结论的基础。但是，通过将看似不太可能的、低熵的宇宙起源纳入到我们的分析中，我们现在明白正确结论应该是：熵会朝着未来的方向增长，因为概率论证完全有效并且在该方向上没有限制；但熵不会朝过去的方向增长，因为这样运用概率将与我们新的附加条件——宇宙开始于低熵而非高熵状态——相冲突。[17] 因此，宇宙诞生时条件对时间之箭的方向非常重要。未来就是熵不断增长的方向。时间之箭——事物这样开始那样结束，而不会那样开始这样结束这个事实——在新生宇宙那高度有序的低熵状态中开始了自己的旅程。[18]

## 未解之谜

早期宇宙为时间之箭设定了方向这个结论美妙而令人满意，但故事还没有结束。重大的谜题仍然没有解开。宇宙开始于高度有序的形态，在接下来的几十亿年间，世间万物慢慢地向着有序度低的方向演化，熵一点点地增加，那么，宇宙是怎样做到这些事的呢？千万别忽略这个问题的重要性。我们曾强调过，从概率的观点来看，你之所以会在晚上10：30看到部分融化的冰，更为可能的原因是杯中水发生了统计学上的偶然事件，而不是之前有一块完整的冰块。对于冰块而言的正确的东西，对于宇宙而言也总是正确的。从概率的角度来说，现在我们在宇宙中所看到的每一样东西，更有可能源于虽然少见但偶尔却会发生的整体无序度的统计偏差；相比之下，从大爆炸所要求的不可思议的高度有序的低熵起点开始，慢慢地演化到现在的高熵状态这种说法，正确的可能性更低。[19]

但是，当我们用概率来考虑问题，将世间万物都想象成由于统计学上的偶然事件才存在于这个世界时，我们会发现自己深陷困境：这种思路让我们开始怀疑物理定律本身。因此我们倾向于反对用统计学上的偶然事件，而更愿意用低熵的大爆炸来解释时间之箭。这样一

来，问题就变成了弄清楚宇宙是怎样从这样一种看似不太可能的、高度有序的形态开始一切的。这才是时间之箭所需要的问题。所有一切最后都归结到宇宙学上。[20]

我们将在第8章到第11章中仔细地讨论宇宙学。首先要注意的是，在我们有关时间的讨论中存在着一系列的缺点：我们讨论过的一切都只基于经典物理。现在我们要来看一下，量子物理会对我们理解时间、追索时间之箭产生哪些影响。

## 第7章　时间与量子

从量子角度洞悉时间的奥秘

当我们思考一些事物，比如时间，比如那些我们置身于其中的事物，比如那些完全融入我们日常生活的事物，比如那些四处弥漫的事物时，其实我们很难——哪怕暂时一下——做到不受通俗语言的影响，我们的思考很难摆脱经验的影响。这些日常经验只能算是经典体系中的经验，会在很高的精确度上符合300多年前牛顿所创立的物理定律体系。但是，在过去的100年间所有的物理学发现中，量子力学无疑是最令人吃惊的，因为它破坏了经典物理学的整个概念体系。

因此，我们很有必要将我们的经典物理经验推广到量子领域，看看那些能够展现量子过程随时间演变时出现奇异特性的实验。在这个背景下，我们将继续上一章的讨论，探寻量子力学描述下的自然界中是否存在时间之箭。我们将得到一个结论，虽然该结论在物理学家中还存在着争议。我们将再一次回到宇宙起源的问题上。

量子论中的过去

在上一章中，概率扮演着核心的角色。但是，正如我一再强调的，概率之所以如此重要完全在于它在实际应用上的便捷以及它所提供的信息的有用性。精确地计算一杯水中的$10^{24}$个$H_2O$分子的运动远

远超越了我们的计算能力；而且，就算我们有这个计算能力，我们又能拿堆积如山的数据怎么办？从$10^{24}$组位置和速度的数据中看出杯中是否出现冰块绝对是一项艰巨的任务。所以我们还不如干脆寻求概率的帮助呢，概率的好处并不仅仅在于我们能够对付得了其中的计算，还在于使用概率方法时我们讨论的是宏观性质——有序还是无序；比如说，是冰还是水——而这正是我们感兴趣之处。但别忘了，我们还没有办法将概率整合进经典物理学的框架中。原则上讲，如果我们准确地知道了事物现在的状况——构成宇宙的每个单独粒子的位置和速度——经典物理学告诉我们可以利用这些信息来预测事物在未来或过去某一特定时刻的状况。理论上，你是否能弄清事物每时每刻的情况——根据经典物理你可以将其称之为过去和未来——取决于你对现在所做观测的精细度。[1]

在本章中，概率将继续扮演着重要角色。但是，因为概率是量子力学中一个不可或缺的因素，它从根本上改变了我们对过去和未来的概念。我们都知道，量子力学的不确定性使我们无法同时知道物体的精确位置和速度。相应地，量子力学预言的只是这样或那样的未来成真的概率。我们当然对这些概率有信心，但它们也只是概率而已，因而预测未来时总是存在不可避免的偶然因素。

在描述过去方面，经典物理和量子力学之间也存在很大的不同。在经典物理学中，为了平等对待所有时刻，我们在描述导致我们所观测到的事物的事件时所用的语言，完全等同于我们在描述观测本身时所用的语言。如果我们在漆黑的夜空看到一颗流星飞过，我们可以讨论它的位置和速度；如果我们想弄明白它是怎样到达这儿的，我们也得搞清楚当它穿过太空飞向地球时的一系列位置和速度。而在量子力学中，一旦我们观测到某物，我们就到了一片净土，在这里，我们对所知道的事情有100%的把握（与该问题有关的仪器精确性及其他类似的问题暂时忽略）。但是，过去——特别是那些“没有被观测到”的过去，在我们，或任何其他人，任何事物进行某一观测之前——存在

于量子不确定所带来的概率王国中。即使我们于此时此刻此地碰巧测量到了一个电子的位置，但在此之前，我们所知道的一切不过是这个电子在这儿或在那儿或在其他任意位置的概率。

而且，正如我们所看到的，并非电子（或者是其他粒子）位于这些可能位置中的一个，只是我们不知道到底是哪个这么简单。[2]实际情况是，所有的位置对电子而言都是有一定意义的，因为每一种可能性——每一种可能的历史——都对我们现在所观测到的结果有贡献。别忘了，在第4章中，我们已经知道可在实验中看到相关证据——电子被迫通过两条缝隙。经典物理学使人们普遍的存有这样的信念：任何事物都有其独一无二的固有历史，所以人们会认为任何一个电子要么从左边的缝隙穿过，要么从右边的缝隙穿过，然后才能到达接收屏。然而，有关过去的这种观点会使我们误入歧途：它预测的结果[如图4.3（a）所示]与实际所发生的情况[如图4.3（b）所示]并不相符。只能借助于通过这两条缝隙的某物的叠加才能解释观测到的干涉图样。

量子力学提供了这样一种解释，但这样做戏剧性地改变了我们对过去——我们对自己观测到的某种事物的由来的描述——的认识。根据量子力学，每个电子的概率波确实穿过了这两条缝隙，正是由于来自每个缝隙的波相互混合，才使得最后的概率波呈现出干涉图样，从而使得电子所落的位置呈现出干涉图样。

与日常经验相比，我们完全不熟悉这种用概率波的混杂来描述电子历史的方式。但是，管他呢，你可能会认为进一步采用这种量子力学描述，会被带到某种更为怪异的可能性前。或许每个单独的电子在到达屏幕之前都会经过两个缝隙，所得到的实验数据不过是两种历史的干涉。也就是说，我们可能会忍不住这样想，来自于双缝的波实际代表的是单个电子的两种可能历史——通过左边的缝隙或右边的缝隙——而且，既然这两列波都对我们从屏幕上观测到的结果有贡献，那么量子力学或许是在告诉我们，每个电子的两种可能历史都对结果有

贡献。

令人惊奇的是，这种奇妙的想法——20世纪最富有创造性的物理学家之一，诺贝尔桂冠获得者理查德·费曼的脑力结晶——提供了一种思考量子力学的完美又可行的方法。根据费曼的想法，如果达到某一给定结果的方式有很多种——比如说，一个电子既可通过左边的缝隙到达探测屏的某一点，又可通过右边的缝隙到达探测屏上的同一点——那么我们就可以认为每一种历史都可以发生，而且是同时发生。费曼证明，每一种情况都对它们共同实现的结果的概率有贡献，如果将这些贡献正确地加起来，结果将与量子力学所预测的总概率一致。

费曼把这种想法称为量子力学的历史求和方法，它告诉我们概率波蕴藏着观测之前的所有过去，而且还告诉我们，量子力学要想沿着经典力学失败之处继续前行，就不得不拓展历史的概念。[3]

去往奥兹国[33]

在另一个版本的双缝实验中，不同历史的干涉更加明显，因为到达探测屏的两种路线被分得更开。用光子来做这个实验比用电子更容易一些，因此，我们改用光子源——激光——来做这个实验，我们将激光射入分束器。分束器由半镶银的镜子制成，就像监视器上用的那种，可以使一半光反射回去而使另一半光通过。初始的单束光分裂成两束——左边的光束和右边的光束，就像双缝实验一样，一束光分成了两束。如图7.1那样，合理地放置完全反射的镜子，两束光被一起反射到下面的探测器上。把光看成一种波，就如麦克斯韦描述的那样，我们期望在探测屏上找到干涉图样。左边和右边的光束距离探测屏上除了中心点以外的所有点的光程都略有不同，因此当左边光束在探测屏上某点形成波峰时，右边光束在该点形成的则可能是波谷、波峰或波峰波谷之间的部分。探测屏会记录下两列波合起来的高度，因此会有独特的干涉图样。

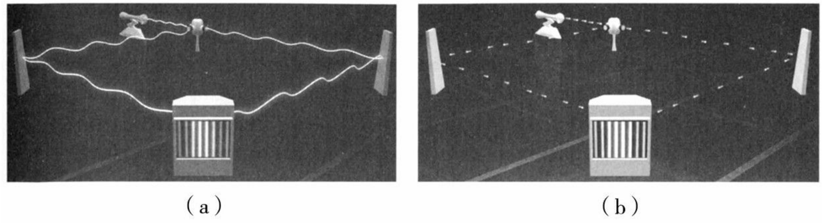


图7.1　（a）：在双缝干涉实验中，激光束分成两股，沿两条路径分别进入探测屏。（b）：将激光调小，使得光子一个一个的出来；一段时间以后，我们还是会看到相干图样。

当我们显著地减弱激光的强度，使其发射出单个光子，比如说每隔几秒发射一个光子时，经典物理和量子物理之间的区别就变得非常明显了。当单独一个光子进入分束器时，经典物理学会告诉我们，它要么穿过去要么被反射回来。经典推理不允许存在一点干涉，因为没有什么可干涉的：从光源射出到达探测屏的只是一个个独立、特殊的光子，一个接一个，有的从左侧过去，有的从右侧过去。但真正实验时（图4.4），记录下来的一个个光子确实产生了如图7.1（b）所示的干涉图样。按照量子力学，这是因为每个探测到的光子可能通过左边或右边的路径到达探测器。因此，我们不得不综合考虑两种历史以确定光子撞击在屏上这点或那点的概率。当每个光子的左边概率波和右边概率波按这种方式组合到一起时，就会通过波的干涉产生概率图样。所以，不像多萝西——当稻草人给她指路去奥兹国时既指左又指右，令她很迷惑——我们所得数据可以完美地被解释为每个光子可以同时通过左右路径到达探测器。

选择

虽然我们在上文只通过几个特殊例子来说明可能历史的组合，但这种思考量子力学的思维方式却具有一般性。经典物理学所描述的现在有一个独一无二的过去，而量子力学的概率波扩大了历史的含义：在费曼的体系里，我们所观测到的现在代表了一种混合——一种特殊的平均——与我们现在所看到的一切相符的所有可能的过去的混合。

在双缝实验和分束器实验中，电子或光子从光源到探测器有两种选择——左边或右边的路径——只有把所有可能的历史组合起来我们才能解释观测到的一切。如果障碍物有3条缝，我们将不得不考虑3种可能的历史；如果有300条缝隙，我们就需要考虑所有可能历史的贡献。现在我们来考虑一种极限情况，如果障碍物上有无数条缝隙——缝隙如此之多以至于障碍物都可以当做不存在了——则根据量子力学，每个电子会踏遍每一条可能的路径以到达探测器上的某一点，只有把与每一种可能历史相关的概率都考虑进去，我们才能解释得到的数据。听起来这或许有点奇怪（确实很奇怪）。但正是这种奇怪的处理过去时间的方法解释了图4.4、图7.1（b），以及每一个探索微观世界的其他实验中的数据。

你可能想知道历史求和这种说法的准确含义到底是什么。电子真的是踏遍了所有可能的路径才撞到探测器上的吗？还是说费曼的说法只是一种能够得到正确答案的巧妙数学设计？这是评价量子实在性本质的关键问题之一，因此我希望我能给出一个明确的答案。但是我做不到。物理学家们常常发现这种把历史求和起来综合考虑的方法非常有用；我在我自己的研究工作中经常使用这种思想，因此我当然觉得它是对的。但是那和说它就是真的还不是一回事。关键在于，量子计算明确地告诉我们电子落在屏幕上这一点或那一点的概率，而这些预测又与数据相符。一旦我们考虑到了理论在预言上的有效性，电子究竟是如何到达屏幕上某点的就不再那么重要了。

当然，你可能会进一步想到，我们也可以解决到底发生了什么这个问题，只要我们改变实验条件，我们也能看到带来了所观测到的现在的各种可能过去的大杂烩。这是一个好建议，但我们也知道还存在另外一个问题。在第4章中，我们知道概率波并不能直接观测到；而费曼把各种历史结合起来的想法也只不过是一种思考概率波的特殊方式，因而它们也没法被直接观测。确实如此。观测不能区分各种历史；相反，观测反映的是所有可能历史的平均。因此，如果你改变了

实验条件，再观测飞行中的电子时，你将会看到每个电子在或这或那的位置穿过额外的探测器；你永远不会看到任何的多重历史。当你用量子力学来解释为什么你会在或这或那的位置看到电子时，答案将与导致中间观测现象出现的所有可能历史的平均有关。但是观测本身只能针对已经求和的历史有关。观测飞行中的电子时，你已经将你所谓历史的概念推后了。量子力学极其狡猾：它解释了你所看到的东西，但又不让你看到解释。

你可能会进一步追问：那么为什么用单独的历史和轨迹描述运动的经典物理学——常识物理学——竟可以解释宇宙？为什么经典物理学在解释和预测每一样物体（从棒球到行星到彗星）的运动时都如此有效？为什么日常生活中就没有证据说明过去会以这种奇特的方式发展到现在？正如我们在第4章中简要介绍并要在稍后更为详尽地探讨的那样，这里的原因在于，与电子之类的粒子相比，棒球、行星和彗星都比较大。在量子力学中，某物越大，就越会偏离平均：所有可能的轨迹确实都对棒球的飞行有贡献，但我们通常看到的棒球轨迹——牛顿定律所预测的那条——比其他路径合起来的贡献还要大很多。对于大个物体而言，经典路径是平均过程中的主导贡献，而且远大于其他贡献之和，因此经典路径才是我们最熟悉的路径。但是，当物体非常小时，像电子、夸克和光子，各种历史不分伯仲，都对平均过程的形成起重要作用。

最后你可能会问：为什么观测和测量的作用如此特别，以至于会迫使所有可能的历史结合到一起，导致单独的一个结果？我们的观测行为又是如何告诉粒子该什么时候将历史求和起来，平均一下并得出一个明确结果的呢？为什么我们人类和我们制作出的机器有这种特殊的力量呢？这特殊吗？又或者，人类的观测行为只不过是更为广义的环境影响的一个子集，我们根本就不特殊？在本章的后半部分，我们将着手讨论这些令人迷惑而又富于争议的问题，因为它们不仅对于量子实在性的本质非常重要，还能为探讨量子力学和时间之箭提供了一

个重要的理论框架。

计算量子力学的平均值需要严格的技术训练。彻底理解平均值是何时何地如何求和起来的，则需借助于物理学家们仍然在努力探索的概念。但关键的一点可以简单地表述为：量子力学是终极的选择舞台：每一种可能的“选择”（从这里到那里时需要做出的抉择）都被包括在与这样或那样的可能结果相关的量子力学概率中。

经典物理和量子物理对待过去的方式完全不同。

修剪历史

以我们所受的经典物理教育去想象一个不可分的物体——电子或光子——同时沿着多条路径运动，是极为不可思议的。即使是我们当中最有自制力的人，也难以抵制偷偷观测一下的诱惑：当电子或光子通过双缝屏幕或分束器时，为什么不偷看一下它们究竟是通过哪条路径到达探测器的呢？在双缝实验中，为什么不把一个小探测器放在每个缝隙前面，以辨别电子到底通过这条缝隙，那条缝隙，还是同时通过这两条缝隙（然后继续前进进入主探测器）？在分束器实验中，为什么不在每条发射路径中放置一个小小的探测器以鉴别光子通过哪条路径，左边的？右边的？还是同时通过这两条（光子继续朝探测器前进）？

答案是你可以插入额外的探测器，但如果你这样做了，你会发现两件事情。第一，你将发现每个电子和每个光子总是会通过一个探测器并且只能通过一个探测器，也就是说，你能确定电子或光子通过了哪条路径，你将发现它总是通过其中一条路径，而不能两条都通过。第二，你将发现主探测器记录的最终数据发生了改变。你看到的不是图3.4（b）和图7.1（b）的干涉图样，而是如图4.3（a）中经典物理学所预测的结果。通过引进新元素——新探测器——你已经在不经意间改变了实验。这种改变规避了你之前要探讨的矛盾——现在你已知

道了每个粒子将通过哪一条路径，这样一来又怎么能和它明显没通过的路径发生干涉呢？之所以会这样，可以用上一节中的讨论来解释。你的新观测会挑选出那些你的最新观测可以探明的历史。这些观测确定了光子会通过哪一条路径，这样我们就只需考虑那些通过这条路径的历史，从而排除了干涉的可能性。

尼尔斯·玻尔喜欢用他的互补原理来总结类似的事情。每个电子，每个光子，事实上所有的事物，都同时具有波动性的一面和粒子性的一面。这些性质具有互补性。只按传统的粒子观点——粒子按独一无二的轨迹移动——来考虑问题并不完备，因为它忽略了干涉图样所展现的波动性一面。[34]但只从波动性的一面来考虑问题也是不完备的，因为这样就忽略了定位电子之类的测量——比如说，通过记录屏幕上的点来定位电子——所展现出来的粒子性特点（图4.4）。完整的描述应当同时把互补性的特点都考虑进来。在任意给定的情况下，你可以通过选择相互作用的方式来使其中一个特点更加明显。如果你让电子从光源发射到未观测的屏幕，其波动性的特点将显现出来，产生干涉。但如果你观测到了电子的路径，你知道它走的是哪条路，你就很难解释干涉性。这时实在性会伸出援手。你的观测直接排除了各种可能的量子历史。它使电子表现得像一个粒子；因为粒子总走这条路或那条路，没有干涉图样，所以没有什么需要解释的。

大自然会做出很奇怪的事情。它总爱打擦边球。但又总是很小心地在致命的逻辑陷阱边迂回而过。

历史的不可期

这些实验非常著名。它们提供了简单而有力的证据证明，掌控着我们的世界的定律是物理学家在20世纪所发现的量子定律，而不是牛顿、麦克斯韦和爱因斯坦所发现的经典定律——这些定律在描述大尺度上的事件时可作为极为有力的近似。我们现在所看到的量子定律挑战了有关过去的传统概念——那些我们未观测到的事件正是造成我们

现在所见到的结果的原因。这种实验的一些简单版本以更为令人惊奇的方式冲击着我们直觉上的对事物随时间演化的看法。

第一个变体是所谓的延迟选择实验，由著名物理学家约翰·惠勒于20世纪80年代提出。这个实验与一个听起来相当怪诞的问题有关：过去取决于未来吗？注意，这里并不是说让我们回到过去，改变过去（这个问题我们将在第15章中阐述）。相反，惠勒实验——人们已经仔细地分析实行过这个实验——探讨的是那些在我们的想象中过去——即便是遥远的过去——可能发生的事件，与那些我们看到的正在发生的事件之间的富有争议性的相互影响。

为了便于理解物理学，想象一下你是位艺术品收藏家，新斯普林菲尔德艺术与美化协会的主席史密瑟先生来观看你所收集的用于拍卖的各种物品。你知道，他真正感兴趣的是《脱衣舞男》，这是一幅你自己不太喜欢的画，但这是你深爱的伯祖父伯恩斯留给你的，因此决定到底要不要卖它是一番情感上的斗争。史密瑟先生来后，你和他谈论着你的收藏，最近的拍卖，最近在大都会博物馆的展览。令人惊奇的是，你得知，许多年前史密瑟先生曾是你伯祖父的得力助手。谈话的最后，你决定放弃《脱衣舞男》：还有许多你想要的其他作品，你必须学会放弃，否则你的收藏将没有焦点。在艺术收藏的世界里，你总在告诉自己，有时更多就是更少。

当你反思这个决定时，你发现事实上在史密瑟先生到来之前你就已经决定要卖掉它。虽然一直以来你都对《脱衣舞男》有种特殊的感情，但你一直在努力避免没有计划、漫无目的的收藏，而且以20世纪末的色情现实主义为主题的收藏几乎被视为只有最有经验的老手才可踏足的收藏禁地。即使你记着在你的客人到来之前你不知道应该怎么办，但从你现在的做法来看，你当时确实已经决定了。这并不是说未来发生的事件影响了过去，而是说你和史密瑟先生的会面，以及接下来做出的愿意卖画的声明表明，你早以某种方式做出了明确的决定，虽然在当时看来你并没决定。就好像是这次会面和你的声明帮助你接

受了这个已经做出的决定，而这个决定只是等待着被发掘出来。未来帮助你知晓过去到底发生了什么。

当然，在这个例子中，未来的事件只是影响了你对过去的观点和阐释，因此这些事件既不令人困惑也不令人惊讶。但是惠勒的延迟选择实验把这种未来和过去之间的心理上的相互作用转移到量子领域，这就变得非常精确而且令人相当吃惊了。实验开始时的设置如图7.1（a），把激光调弱使其一次只发射一个光子，如图7.1（b）那样，同时在分束器旁加放一台新的光子探测器。如果关掉新的探测器[图7.2（b）]，则我们回到了初始实验条件，接收屏上就会出现光子的干涉图样。但如果打开新的探测器[图7.2（a）]，它就会告诉我们每个光子经过哪一条路径：如果它探测到一个光子，那么光子走的就是这条路径；如果它没探测到光子，光子走的就是另一条路径。这种所谓的“路径选择”信息促使光子表现得像粒子一样，因此不再产生波的干涉图样。

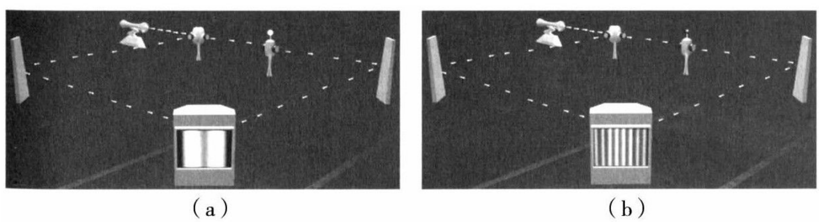



图7.2　（a）：打开“路径选择”探测器，我们就破坏了相干图样。（b）：当新的探测器关掉的时候，我们就回到了图7.1的情形，相干图样重新出现。

现在我们来改变一下实验条件，按惠勒的方式，沿着两条路径之一移动新的光子探测器。从原理上讲，这两条路径可以无限长，因此新的探测器可以距离分束器相当长的距离。如果关掉新的探测器，我们就又处于通常的情况下，屏幕上将全是干涉图样。如果打开它，它将提供路径选择信息，进而排除干涉图样的存在。

新的诡异之处在于这样一个事实：路径选择的测量发生在光子在

分束器中不得不“决定”是像波一样同时经过两条路径还是像粒子一样只经过一条路径很长时间以后。当光子经过分束器时，它无法“知道”新的探测器是关着还是开着——事实上，可以在光子经过分束器之后再设定探测器的开关。在探测器关掉的情况，光子的量子波最好分裂，同时沿着两条路径传播，这样一来，两列波的叠加就会产生干涉图样。但是，如果新的探测器一直开着——又或是在光子完全经过探测器后才打开——那光子就会遭遇身份危机：光子本来已经通过两条路都走确定了自己具有波动性，但现在在做出选择之后，它“意识”到它需要成为一个粒子，沿着一条路径运动，并且只沿着一条路径运动。

但不管怎样，光子总不会犯错。不管探测器什么时候打开——即使迟至某个光子通过分束器后再打开探测器——光子仍然像个粒子那样运动。我们发现它总是通过单独一条路径飞向屏幕（如果我们在两条路径都放置有光子探测器，那么激光器发射出去的每个光子将只被一个探测器观测到，而不是两个都能观测到）；最后的数据将不会展现任何干涉性。无论什么时候关掉探测器——再次，即便是在每个光子都通过分束器后才做出决定——光子也会表现出波动性，产生显著的干涉图样，表明它们通过的是两条路径。似乎光子会根据未来新探测器是打开还是关闭来调整它们过去的行为；似乎光子可以预先得知它们在下面的路途中会遇到何种实验条件并提前做出相应的行为一样。似乎一段可靠确定的历史只有在其所导向的未来完全定下来之后才会变得清楚。[4]

这与你是否决定要卖出《脱衣舞男》的经历有一定的相似性。在遇到史密瑟先生之前，你正处在一个模糊、还未决定、既愿意卖画也不愿意卖画的混合状态。但是，一起讨论过艺术世界，并且得知史密瑟先生对你伯祖父的感情之后，卖画的想法就在你的头脑中定型了。这次谈话使你下定决心，这样的决定使这段决定的历史从先前的不确定中明晰起来。反思过去，就好像这个决定早就做出一样。但如果你

和史密瑟先生相处不是十分愉快的话，如果他没有获取你的信任让你觉得《脱衣舞男》并不会在他手中辱没，你或许就会觉得不卖出去挺好的。在这种情况下，你可能会觉得事实上很久以前你就决定不卖这幅画——不管卖出这幅画多么的明智，但在你内心深处感情的维系让你对这幅画无法释怀。事实上，过去一点儿也没有改变。只是现在的不同使你对过去描述有所不同。

在心理学领域，重写或重新诠释过去是很常见的事情。我们常常通过现在的经历获知过去的故事。但在物理学领域——一个我们通常认为是很客观的领域——未来的偶然事件竟会使过去变得不同则令我们感到头晕。为了使人们更加头晕目眩，惠勒想出了宇宙学版本的延迟选择实验，光源不是实验室中的激光，而是宇宙深处强有力的类星体。分束器也不是实验室的那种，而是居间星系，它们的引力可以像透镜那样聚焦经过的光子，指引它们向地球运动，如图7.3那样。虽然没有人做过这个实验，但从原理上讲，如果收集到足够多的来自类星体的光子，它们就应该可以在长期曝光的相片底板上产生干涉图样，就像在实验室里的分束器实验一样。但是，如果我们把一个额外的光子探测器放在某条路径的末端，它就会为光子提供路径选择信息，从而破坏干涉图样。

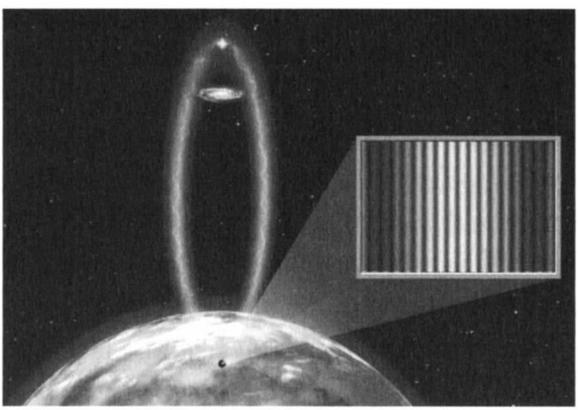


图7.3　来自于远方的类星体的光，会被中间的星系劈裂及会聚，这样的光，至少在理论上会产生相干图样。如果有另外一个可以确定每一个光子所走路径的探测器开着的话，则光子将不会再产生干涉图样。

这个版本的实验令人吃惊之处在于，从我们的角度来看，这些光子来自几十亿光年外。到底是像粒子那样沿着一条路径运动，还是像波那样沿着两条路径运动，它们的这个决定看来早在探测器、我们人类甚至是地球存在以前就已经做出来了。但是，几十亿年后，探测器被制造出来，安装在光子到达地球的路径上并扭开开关。这些近期的行为不知为何确保了被观测的光子呈粒子样运动。它们表现得就好像它们一直都精确地沿着朝向地球的某条路径运动。但是，如果几分钟后，我们关掉探测器，接着到达相片底板的光子就会造成干涉图样，就好像几十亿年来，它们一直与其幽灵般的同伴一道飞向地球一样，只不过它们的同伴在飞越居间星系时会走与它们相反的路径。

我们在21世纪打开或关掉探测器会对几十亿年前的光子运动产生影响吗？当然不会。量子力学并不否定已经发生的过去。问题源于量子中的过去概念不同于经典直觉中的过去概念。我们的经典教育使我们长时间以来一直说某个光子做过这个，做过那个。但在量子世界，也就是我们的世界中，这种思维强加给光子一种限制过度的实在性。就像我们所看到的，在量子力学中，正常态是一种不确定的、模糊的、混乱的、千丝万缕的实在性，只有进行一定的观测时，它们才会清楚的变成一种更为大家熟悉的、明确的实在性。光子并不是在几十亿年前就决定了到底是按某条路径绕星系运动，还是同时沿两条路径运动。相反，几十亿年来它一直处在量子的正常态——各种可能性的混合。

这种观测将不熟悉的量子实在性与日常的经典经验联系起来。我们今天所做的观测使量子历史的某一缕在我们探讨过去时变得重要起来。在这种意义上来讲，虽然从过去到现在的量子演化不受我们现在所做的任何事情影响，但是，我们所讲的有关过去的故事则会留有今天行为的痕迹。如果我们在光射向屏幕的途中插入光子探测器，那

么，我们有关过去的故事就将包括每个光子走的到底是哪一条路径这样的内容；通过插入光子探测器，我们保证到底是哪条路径这一信息是我们的故事中重要而又确定的细节。但是，如果我们不插入光子探测器，我们有关过去的故事就会全然不同。没有光子探测器，我们就不能说清光子走的到底是哪一条路径；没有光子探测器，就无法获知到底是哪一条路径。两个故事都是正确的。两个故事都很有趣。两者所描述的只是不同的情形而已。

因此，今天的观测帮我们讲完了一个有关开始于昨天、前天，甚至是10亿年前的过程的故事。今天的观测勾勒出的细节，我们可以而且必须包括在今日之对过去的描述中。

抹掉过去

需要特别注意的是，在这些实验中，过去不会被今天的行为以任何形式改变，实验的任何修正都无法完成这样一个难以企及的目标。这就提出了一个问题：如果你不能改变已经发生的事情，那么你能做哪些事情来消除其对现在的影响呢？从某种程度上讲，有时这种幻想可以成真。一名棒球手，在第9局最后己方已经两人出局的情况下，错失了一次普通的击球，使对方成功将自己一方封在一垒；不过，只要能够将下个投手掷出的球打好，他就可以挽回自己的错误。当然，这样的例子毫无神秘之处。只有当过去的某个事件干脆利落的除掉了未来另一件事情发生的可能性（比如说，那名球员击球后被对方直接接杀就意味着他们队完了），而我们随后又得知那件不可能的事情却发生了时，我们才会意识到有些东西非常奇怪。玛兰·斯考利和凯·德鲁尔于1982年首次提出的量子橡皮，就暗示我们量子力学存在这种奇怪现象。

量子橡皮实验的简单版本利用的是双缝实验的装置，只不过要以如下方式稍作修改。每个缝隙前面都放置一个标记装置，它会为每一个经过的光子做记号，这样一来，稍后只要查验光子，你就可以知晓

它所通过的到底是哪一条缝隙。如何标记一个光子——你该如何在从左边缝隙通过的光子身上标一个“L”，在右边缝隙通过的光子身上标一个“R”——的确是一个好问题，不过细节并不重要。粗略地讲，可以用这样的方法标记，让光子自由地通过某个缝隙，然后迫使其自旋指向某个特殊方向。如果左右缝隙前的装置能使光子的自旋指向特定但又不同的方向，那么我们就可以借助于一台更加精密的接收屏——这个新的接收屏不仅可以标记光子落在屏上何处，还可以记录下光子的自旋指向——来搞清楚光子到底通过的是哪一条缝隙。

实施这个带标记的双缝实验时，光子并没有形成如图7.4（a）所示的干涉图样。现在我们应该已经很熟悉这里的解释了：新的标记装置会获得有关哪一条路径的信息，而哪一条路径信息又能够选定或这或那的历史；实验数据会告诉我们，某个光子通过的到底是左边的缝隙还是右边的缝隙。如果没有经过左边缝隙和经过右边缝隙的轨迹的组合，就不会有概率波的叠加，因而就不会产生干涉图样。

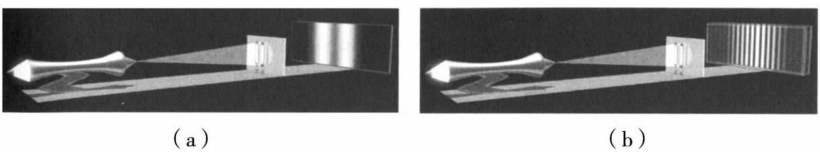



图7.4　图（a）在量子橡皮实验中，双缝前面的设备用来标记光子，以便弄清每一个光子穿过的究竟是哪一条缝隙。我们在图（a）中看到，这个判断路径的信息破坏了干涉图样。在图（b）中，另一个设备，用于抹掉光子上的标记的设备，被放置在探测屏前，由于判断路径的信息被抹掉了，干涉图样又重新出现了。

现在，我们来看看斯考利和德鲁尔的想法。在光子撞击接收屏之前，如果你把标签装置对光子所做的标记擦除，从而消除了获知光子通过哪条缝隙的可能性，那又会怎样？这样一来，即使从理论上讲，也没有办法从探测到的光子中获取哪条路径的信息，这会使两种历史发生相互作用，从而形成干涉图样吗？注意这种“取消”过去可比棒球手在第9局最后的神奇接球厉害多了。按下标签装置的开关时，我们可以想象每个光子都像粒子般运动，穿过左边的缝隙或右边的缝隙。

不管通过什么方法，在光子撞上屏幕之前，我们将其上所记录的有关通过哪一条缝隙的信息擦除掉了；但是，似乎对于形成干涉图样而言已太晚。在干涉中，光子呈现波动性。它必须同时经过两条缝隙，这样它才能在到达探测屏的过程中相互混合。但我们起初对光子所做的标记似乎保证它会像粒子一样运动，要么经过左边的缝隙，要么经过右边的缝隙，从而使干涉过程不会发生。

在雷蒙德·齐奥、保罗·奎特和埃弗雷姆·斯特恩伯格做的实验中，实验装置如图7.4所示，有一个新的擦除装置插在探测屏之前。虽然细节并不重要，但还是简要介绍一下，不管光子是从左边的缝隙还是右边的缝隙进入，擦除装置都会使其自旋指向同一个固定方向。这样一来，通过测量自旋就不会获得任何信息，没法发现光子通过的是哪条缝隙，所标记的哪一条路径信息被擦除了。神奇的是，擦除之后，屏幕探测到的光子确实产生了干涉图样。当擦除装置被置于接收屏之前时，它消除了——擦除了——光子通过双缝时被标记所带来的影响。就像延迟选择实验中的情形一样，理论上，这种擦除可以在其所要干扰的事件发生的几十亿年后才进行，即使这样也会有效消除过去，甚至是久远过去的影响。

我们怎样来理解其中的意义呢？这个嘛，要记住这些数据与量子力学的理论预言符合得非常完美。斯考利和德鲁尔之所以提出这个实验是因为他们所做的量子力学计算使他们确信一定会发生这样的事。确实就是这样。因此，就像有关量子力学的一般问题一样，谜团并不会使理论与实验相矛盾。这样的实验只会使理论——得到了实验验证的理论——与我们对时间和实在性的直觉相违背。还需要知道的是，如果你在每条缝隙前面放一台光子探测器，探测器就会确定地告诉我们光子通过的到底是左边的缝隙还是右边的缝隙，这样确定的信息没法被擦除，因此也就没有办法重现干涉图样。但标记装置是不一样的，因为它们所提供的只是获得有关哪一条路径信息的可能性——而这种可能性是可以被擦除的。简单地讲，标记装置对经过的光子动了

点手脚，光子仍然通过两条路径，但标记装置使光子概率波的左边部分变得比右边部分模糊，或者使光子概率波的右边部分变得比左边部分模糊。相应地，本应从每条缝隙中按顺序正常出现的波峰波谷——如图4.2（b）——也会变模糊，因此探测屏上就不会形成干涉图样。关键在于，要认识左边的波和右边的波都还存在。擦除装置之所以会起作用是因为它重新聚焦了波。就像一副眼镜一样，它会抵消模糊，使两列波重新聚焦，从而得以再次形成干涉图样。似乎在标记装置的作用下，干涉图样虽然从视野中消失了，但却耐心地守候在那里，等待着某人或某物来拯救它。

或许这种解释使量子橡皮不那么神秘，但这里就是终点了——量子橡皮实验令人惊异的变异版对传统意义上的时间和空间概念构成了更为猛烈的挑战。

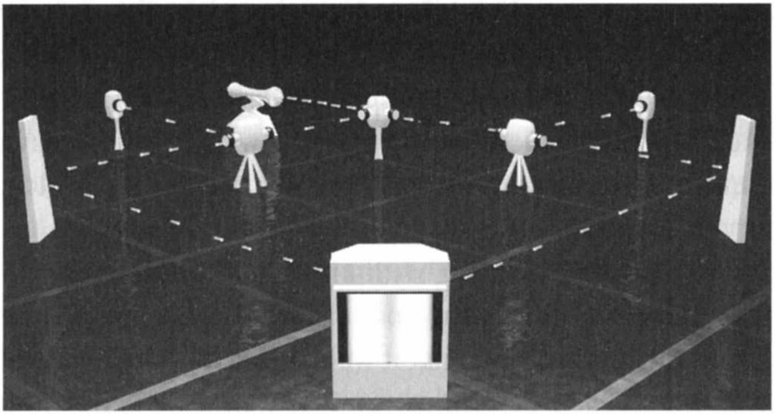
塑造过去[35]

这个实验，延迟选择的量子橡皮擦，也是斯考利和德鲁尔提出的。首先要对图7.1所示的分束器实验加以改进，插入两个所谓的降频转换器，一边一个。降频转换器是这样一种设备，输入一个光子它就能输出两个光子，而每个光子的能量都是原始光子能量的一半（“降频”）。其中一个光子（被叫做信号光子）直接沿着原始光子飞向探测屏的路径运动。同时，降频转换器产生的另一个光子（被叫做闲频光子）则沿不同方向发射出去，如图7.5（a）所示。每次做这个实验时，我们通过观测降频转换器发射出来的闲频光子伴所走过的路径，就可以确定信号光子走的是哪条路径。又一次，获知信号光子走哪条路径的能力——即使是完全间接的，因为我们与任何信号光子之间没有一丝相互作用——阻碍了干涉图样的形成。

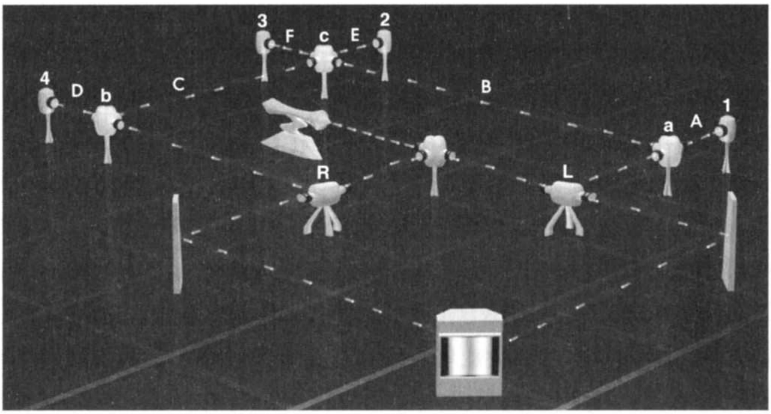
现在我们来看一下更加诡异的部分。要是我们改变实验设置，使我们无法获知某闲频光子到底来自于哪一个降频转换器，会怎样呢？也就是说，如果我们擦除了闲频光子所带有的哪一条路径信息，又会

怎样呢？令人惊奇的事情发生了：即使我们并没有直接对信号光子做什么，通过擦除闲频光子所带有的哪一条路径信息，我们又可以观测到信号光子所形成的干涉图样。让我来告诉你这个过程是怎样发生的，它实在太神奇了。

看一下图7.5（b），它包含了所有实质性的信息。但不要害怕。它实际上比看上去简单些，我们现在就按易于处理的步骤看过一遍。图7.5（b）中的设置不同于图7.5（a）中的设置，其区别在于如何探测发射出来的闲频光子。图7.5（a）中，我们可以直接探测到它们，因此很快就可以确定每个光子是从哪个降频转换器发射出来的——也就是说，特定的信号光子走的是哪条路径。在新的实验中，每个闲频光子都要走一个迷宫，从而使我们没法定出信号路径。比如说，想象一下有一个闲频光子从标着“L”的降频转换器发出。这个光子并没有立即进入探测器[如图7.5（a）所示]，而是被送到分束器（标记为“a”）中，这样它就有50%的概率沿着标为“A”的路径运动，50%的概率沿着标为“B”的路径运动。如果光子沿着A路径向前运动，它就会进入一个光子探测器（标记为“1”），并且会被恰当的记录下来。但如果闲频光子沿着B路径向前运动，那它将继续经历这一切。它将向另一个分束器（标记为“c”）运动，并且有50%的概率沿着E路径运动到达标记为“2”的探测器，有50%的概率沿着F路径运动到达标记为“3”的探测器。现在——跟上我，关键之处要到了——相同的论证也可以应用于标记为“R”的另一个降频转换器发出的闲频光子，如果这个闲频光子沿着D路径运动，它将被探测器4记录；如果它沿着C路径运动，那么根据其通过分束器c后所走的路径的不同，它将被探测器3或探测器2探测到。



（a）




（b）

图7.5　（a）：添加了降频转换器的双缝实验，不会带来干涉图样，因为慢光子会带来路径判断信息。（b）：如果慢光子没有被直接探测到，而是被送进了图中的迷宫，那么就可以从数据中抽取出干涉图样探测器2或探测器3，探测到的慢光子不会导致判断路径信息，因而其信号光子还会带来干涉图样。

现在我们来看看为什么要增加这些复杂性。注意，如果一个闲频光子被探测器1探测到，我们就会得知相应的信号光子沿着左边的路径运动，因为对于从降频探测器R发出的闲频光子而言，没有其他路径可以到达这个探测器。类似地，如果一个闲频光子被探测器4探测到，我们就可以知道它的信号光子伴沿着右边的路径运动。但如果一个闲频光子到达探测器2，我们就不知道它的信号光子伴沿着哪一条

路径运动了，因为它有50%的可能性从降频转换器L发出，沿着路径B-E运动，也有50%的可能性从降频发射器R发出，沿着路径C-E运动。类似地，如果探测器3探测到一个闲频光子，该光子既有可能是从降频发射器L发出沿着路径B-F运动，也有可能是从降频发射器R发出沿着路径C-F运动。因此，如果闲频光子被探测器1或探测器4探测到，我们就可以推测出其相应的信号光子的哪一条路径信息，但如果闲频光子是被探测器2或探测器3探测到，相应的信号光子的哪一条路径信息就被擦除了。

是不是哪一条路径信息的擦除——即使我们并没有直接对信号光子做什么——就意味着干涉效应会重现？确实是这样——但与信号光子相应的闲频光子到达的必须是探测器2或探测器3。也就是说，屏幕上信号光子撞击位置总体上看来与图7.5（a）所示的数据类似，并没有一丁点干涉图样的痕迹，就好像光子走的是这条或那条路。但如果我们把注意力集中到数据点的子集上——比如说，那些其闲频光子伴进入探测器2的信号光子——那么这些子集中的点将形成干涉图样！这些信号光子——其相应的闲频光子碰巧没有提供关于它们所走路径的任何信息——表现得就像它们沿着两条路径运动一样！如果我们可以将设备连接起来从而实现这样的功能：当信号光子相应的闲频光子被探测器2探测到时，屏幕上就显示一个红点来表示信号光子的位置；当信号光子相应的闲频光子被其他探测器探测到时，屏幕上就会显示绿点来表示信号光子的位置，那么，每个人都会看到红点所组成的明暗相间的条带——干涉图样，除非他是色盲。将这里的探测器2替换成探测器3，也有相同的结果。但那些其闲频光子伴进入的是探测器1或探测器4的信号光子，则不会产生这种干涉图样，因为从这些闲频光子中我们可以知道相应的信号光子的哪一条路径信息。

这些结果——已经得到了实验的证实[5]——光彩夺目：由于使用了有可能提供哪一条路径信息的降频转换器，我们无法得到干涉图样，如图7.5（a）所示。没有干涉图样，我们将自然得出结论每个光

子沿左边路径或右边路径运动。但我们现在了解到这个结论下得太仓促了。小心地消除某些闲频光子所带有的潜在的哪一条路径信息，我们诱使数据产生干涉图样，而这意味着有一部分光子实际上走的是两条路径。

还需要注意的是，所有结果中最令人惊讶的是：3台额外的分束器和4台闲频光子探测器可以在实验室的另一边，甚至是宇宙的另一边，因为在我们的讨论中，没有任何东西取决于这些装置接收到闲频光子是在信号光子撞击屏幕之前还是之后。想象一下，这些装置相距非常远，明确起见，比如说有10光年之远，想想这意味着什么。今天你做了如图7.5（b）中的实验，连续记录一大批光子碰撞的位置，结果发现没有任何干涉的痕迹。如果有人让你解释数据，你可能会说由于闲频光子会暴露路径信息，因此每个信号光子明确地沿着左边或右边的路径运动，从而消除了干涉的可能性。但如上而言，这个结论下得也有点过早，这是对于过去的一种完全不成熟的描述。

你看，10年以后，4个光子探测器将会接收到——一个接一个——闲频光子。如果你接下来知晓哪个闲频光子被探测器2探到（比如说，第1个，第7个，第8个，第20个……闲频光子），那么你回过头去查看早年收集的数据并突出加亮相应的信号光子（比如说，第1个，第7个，第8个，第20个……信号光子）在屏幕上的位置的话，你将发现加亮的数据点形成干涉图样，从而获知那些信号光子走过的是两条路径。而且，如果9年前，也就是你收集信号光子数据的364天后，一个家伙开玩笑拿走了分束器a和b从而破坏了实验——这样就保证了第二天闲频光子再到达时，就只能到达探测器1或探测器4，进而保留了所有的哪一条路径信息——那么，当你知道这件事时，将得出结论说每个信号光子要么沿左边路径运动或要么沿右边路径运动，因此无法从信号光子的数据中得到干涉图样。因此，就像上述讨论中着力强调的那样，你用来解释信号光子数据的故事强烈地依赖于收集数据10年后所做的测量。

让我再来强调一次，未来进行的测量并不会改变你现在所做实验的任何方面；未来的测量不会以任何方式改变你现在所收集的数据。但是，当你接着描述今天所发生的一切时，未来的测量确实会对你所说的细节有影响。在你获得闲频光子的测量结果之前，有关给定信号光子的哪一条路径信息，你真的什么都说不出来。但是，只要你得到了测量结果，你就可以得出结论说，我们成功地利用了信号光子的闲频光子伴，得到了信号光子的路径信息，进而确定信号光子许多年前的运动路径是左边还是右边？同时你也可以得出结论说，如果通过信号光子相应的闲频光子伴所得到的路径信息被擦除，我们就不能说信号光子在许多年前走过的是这条或是那条路径（你可以利用新得到的闲频光子数据来发掘出信号光子数据中隐藏的干涉图样，从而相信这一结论）。因此我们可以看出，未来帮助你讲述过去的故事。

这些实验强烈地冲击着我们传统的空间和时间概念。如果要描述某物，在我们描述某事时，那些发生于其后很久的事件和那些距离其很远的事件非常重要。从经典物理——常识——的角度看，这种说法十分荒谬。当然，这就是问题：将经典物理的思维应用于量子宇宙是一种错误。我们从爱因斯坦波多斯基罗森的讨论中学到，量子力学并不具有空间上的定域性。如果你完全理解了那一课——就其本身而言很难理解——这些与跨越了空间和时间的量子纠缠有关的实验，或许看起来就没那么古怪了。但从日常经验的标准看，它们确实古怪。

量子力学和经验

我记得在第一次得知这些实验后的几天中，我十分高兴。我感到自己触及了实在性隐藏起来的一面。通常的经验——世俗、普通的日常活动——突然成为经典物理之假想的一部分，隐藏在量子世界的真实本质背后。突然间，日常生活的世界看起来就像不真实的魔术，哄骗它的观众相信普通的为人所熟知的空间和时间概念，而量子实在性令人吃惊的真相则藏于大自然的妙手之下。

最近一些年来，物理学家们花了很大的力气试图解释大自然的诡计——弄清楚量子物理的基本定律如何幻化成在解释日常经验上如此成功的经典物理定律——从本质上来讲，就是要搞明白当原子和亚原子联合起来成为宏观物体时，它们如何将其魔法般的奇异性隐蔽起来。研究还未到头，但有些问题已经被弄清楚。现在，让我们从量子力学的角度来探讨一下与时间之箭相关的某些特殊问题。

经典物理学基于17世纪晚期牛顿发现的方程；电磁学基于19世纪晚期麦克斯韦发现的方程；狭义相对论基于爱因斯坦1905年发现的方程，而广义相对论则基于他于1915发现的方程。所有的这些方程都有一个共性，它们都忽略了时间之箭的方向问题，认为过去和未来是完全对称的。在他们的方程中无法区分过去和未来，过去和未来是被同等对待的。

量子力学基于欧文·薛定谔于1912年发现的方程。[6]你不需了解有关这个方程的其他任何东西，除了下面这个事实：该方程把某一时刻的量子力学概率波波形——如图4.5——当做输入，然后据此来确定在其他更早或更晚时刻概率波的波形。如果概率波与某个粒子有关，比如说与电子，你就可以用它来预测在某一特定时刻某一特定位置发现电子的概率。像牛顿、麦克斯韦和爱因斯坦的经典物理定律一样，薛定谔的量子定律平等对待过去和未来。一场展现概率波开始于此而结束于彼的“电影”也可以反过来放映——概率波可以开始于彼而结束于此——没有方法可以鉴别哪种演化是正确的，哪种演化是错误的。对于薛定谔方程来说，这两种情况都是同等有效的解。这两种情况都可以代表事物演化的合理方式。[7]

当然，现在提到的“电影”是完全不同于上一章中分析网球运动或鸡蛋摔碎的电影。概率波并不是我们直接看到的事物；并没有摄像机能捕捉到电影中的概率波。相反，我们可以用数学方程式来描述概率波，在我们的脑海中，我们可以想象一下最简单的概率波形状，如图4.5和图4.6所示。我们了解概率波的唯一途径只能是间接的，即通过

测量物理过程来实现。

也就是说，正如在第4章和上述实验中反复强调的那样，标准的量子力学公式用两个截然不同的阶段来描述现象的演变。在第一阶段，一个诸如电子之类的物体的概率波——或者用这个领域中更为准确的语言说，波函数——会根据薛定谔所发现的方程演化。这个方程确保了波函数的形状平稳渐进的变化，就像水波从湖水的一边运动到另一边时的波形变化一样。[36]在第二阶段的标准描述中，我们通过测量电子的位置而与可观测的实在性发生联系，当我们这样做时，波函数的形状突然改变。电子的波函数不像我们平常所熟悉的水波、声波之类：当我们测量电子的位置时，其波函数会突然变得尖锐，如图4.7所示，在无法测量到电子的位置会发生塌缩变为零，而在能测量到电子的位置则会是百分之百的概率。

第一阶段——波函数随薛定谔方程的演化——从数学上看是非常严格而清晰的，可以完全被物理学界接受。第二阶段——关于测量时波函数的塌缩——却完全相反，在过去80年间，往好的方面讲，我们可以说它使物理学家们感到迷惑，往坏的方面讲，我们可以说它带来的麻烦、谜题以及潜在的矛盾消耗了很多物理学家的职业生涯。就像在第4章末提到的那样，困难之处在于，根据薛定谔方程，波函数并不会塌缩。波函数的塌缩是一种附加物。它是在薛定谔发现方程之后，试图解释实验学家们实际看到的现象时引进的。原始的、不塌缩的波函数使我们产生这样一种奇怪的想法：粒子既在这里又在那里，但实验者从没观测到这样的事。他们总是发现粒子明确的处于某个位置；他们没有看到粒子一部分在这儿，另一部分在那儿；测量仪器上的指针不会既指这个值又指另一个值。

当然，同样的道理也适用于我们对周围世界的观察。我们从不曾看到一把椅子既在这里又在那里；我们从未看到月亮既在这部分夜空，又在那部分夜空；我们从未看到一只猫既死了又活着。只要假定测量行为可以诱导波函数放弃量子不确定状态并引领很多可能性中的

某一种（粒子在这儿或在那儿）成为现实，则波函数塌缩的概念就能与我们的经验相一致。

量子测量之谜

但是，实验人员的测量是如何造成波函数塌缩的呢？而实际上，波函数塌缩真的会发生吗？如果发生的话，微观水平上究竟发生了什么呢？所有的测量都会造成波函数塌缩吗？波函数塌缩何时发生，持续多久？既然根据薛定谔方程波函数不会塌缩，那么在量子演化的第二阶段是什么方程代替了薛定谔方程呢？新方程又是如何废黜薛定谔方程，篡夺了其在量子过程中的中坚地位的呢？就我们在这里所关心的时间之箭而言，既然主宰第一阶段的薛定谔方程在区别时间向前和向后上没有多大意义，那么第二阶段的方程是否为测量前后的时间引入了不对称性呢？也就是说，量子力学，包括其通过测量和观测而与日常世界之间建立的结合点，是否为物理学的基本定律引入了时间之箭呢？毕竟，我们先前讨论过量子力学对过去的态度不同于经典物理学，这里所谓的过去指的是某种观测或测量发生之前。通过阶段二的波函数塌缩得以具体化的测量，是否能在过去和未来之间，测量前后之间，建立时间上的不对称性呢？

这些问题还没有得到完全解决，仍然存在着争议。但几十年来，量子理论预言能力几乎没有受这个问题的影响。即使第二阶段仍保持着其神秘性，量子理论的这种阶段一阶段二体系，仍可以预言这种或那种测量结果出现的概率。一次又一次的重复某一实验，弄清或这或那的结果出现的频率，就可以验证理论所给出的预言。这种方法在实验上取得的巨大成功远远超过了由于不能说清第二阶段发生了什么而有的不满意。

但不满意总是有的。这并不是简单地说波函数坍塌的某些细节还没有搞清楚。所谓的量子测量问题，恰如其名，是一个有关量子力学局限性和普适性的问题。这一点很容易看出。阶段一阶段二方法在被

观测者（比如说电子、光子或原子）和进行观测的实验者之间造成了一条鸿沟。在实验者观测之前，波函数随薛定谔方程快乐温和的演化着。之后，实验者开始着手测量，游戏规则突然就变了。薛定谔方程被放到一边，转而由第二阶段的坍塌接手。但是，既然组成实验者及其所用仪器的原子、质子和电子与实验者要研究的原子、质子和电子没有什么不同，那么究竟为什么量子力学会区别对待它们？如果量子力学是一个普适理论，可以毫无限制的应用于一切事物，那它就应当以平等的方式对待被观测者和观测者。

尼尔斯·玻尔不同意这种意见。他认为实验者及其实验仪器不同于基本粒子。虽然它们是由相同的粒子组成，但它们都是基本粒子的“大”集合，因而由经典物理学定律支配。单个原子和亚原子粒子所构成的微观世界与我们所熟悉的人类及其仪器所构成的宏观世界之间，由于大小不同而造成了规则的不同。提出这种界限的动机十分清楚：根据量子力学，微小粒子会既位于这里又位于那里；但对于大的世界，我们日常生活的世界而言，这种事情不复存在。但确切的边界在哪里呢？而且，重要的是，当在日常的宏观世界遭遇原子的微观世界时，这两套规则又是如何衔接的？玻尔认为这些问题已经超出了他或其他人可以回答的范畴。而且，因为不回答相关问题，理论也可以进行精确预言，所以在相当长的一段时间内，这些问题都不在物理学家们亟待解决的关键问题之列。

但为了完全理解量子力学，完全弄清它所说的实在性，了解其在为时间之箭设定方向上所起的作用，我们必须抓住量子测量问题。

在接下来的两小节中，我们将探讨最有希望解决这个问题的一些尝试。要是你在任何时刻都想直奔最后一节量子力学和时间之箭，那么我可以为你简要的归纳下面两小节的内容：通过一些富于创造性的工作，人们已经取得了一些量子测量问题方面的重大进展，但彻底的解决之道还没有找到。许多人认为这个问题是我们的量子定律中最重要的单独缺陷。

## 实在性和量子测量问题

这些年来，人们提出了许多解决量子测量问题的办法。具有讽刺意味的是，虽然这些方法有不同的实在性概念——某些方法之间差别非常之大，但当涉及一名研究者在最普通的实验中会测得什么时，各种方法会彼此符合。表面上它们演的是同一出戏，但是瞟一眼后台你就会知道，它们背后的机理全然不同。

谈到娱乐消遣时，你大概没兴趣知道后台发生了什么；你所感兴趣的只是展现出来的结果。当谈及理解整个宇宙时，你会急不可耐地扯下所有的窗帘，打开所有的门，完完全全地暴露实在性的内在机理。玻尔认为这种急不可耐毫无基础且具有误导性。在他看来，实在性就是一场演出。就像斯波尔丁·格雷[37]的独白一样，实验学家的测量就是全部表演。没有其他什么东西了。按照玻尔的想法，没有什么后台。试图分析波函数何时如何以及为什么放弃所有可能性只留下一种并在测量设备上留下确定的数值会使人错过要害之处。所测得的数值本身才是值得关注的一切。

几十年来，这种观点一直占据主导地位。但是，虽然它可以缓解量子力学带来的思想斗争，人们还是忍不住会想，量子力学神奇的预言能力意味着它非常接近隐藏在宇宙表面规律之下的实在性。人们会忍不住想要更进一步，弄清量子力学是怎样与日常经验联系起来的——量子力学是怎样在波函数与观测之间架起一座桥梁的？观测背后的隐秘实在性究竟是什么？这些年来，许多研究者接受过这种挑战，下面我们就来看看他们提出的一些想法。

有一种想法，其历史根源可以追溯到海森堡，是要放弃将波函数作为量子实在性的客观性质的观点，转而将其视为我们所了解的实在性的一种化身。在我们进行测量之前，我们并不知道电子在什么位置，这种观点提出，电子波函数将电子位置描述为有许多种可能性这件事反映的是我们对电子位置的无知。但在我们测量电子位置的一刹

那，我们对其所在何处的认识突然改变了：理论上讲，我们现在准确地知道了它的位置（根据测不准原理，如果我们知道它的位置，我们就完全无法知晓其速度，但这并非我们现在所要讨论的问题）。根据该观点，我们认知上的突然变化，反映在电子波函数的突然变化上：波函数突然塌缩并呈现出图4.7所示的波峰形状，这就意味着我们知道了电子的确切位置。从这一点上来讲，波函数的突然塌缩也没有那么令人惊讶：当我们知道一些新东西时我们所体验到的认知上的突然改变也无非如此。

惠勒的学生休·埃弗雷特在1957年提出了另一种想法，在他的方案中根本没有波函数塌缩的概念。相反，波函数中所含有的每一种可能结果都有可能发生；只不过每一种结果都发生在各自的宇宙中。这种想法，就是所谓的多世界诠释，“宇宙”的概念被扩充为无穷多个“平行宇宙”——我们宇宙的无穷多个版本——这样一来量子力学预言的任何东西都有可能发生，即使只有很小的可能性，也有可能在某一个版本的宇宙中真正发生。如果波函数说一个电子既有可能在这儿，也有可能在那儿，或是在某个遥远的位置，那么就会有一个电子在这儿的宇宙版本；而在另一个宇宙版本中，你会发现它在那儿；在第3个宇宙版本中，你会发现电子在遥远的那个位置。我们每个人从这一秒到下一秒所做的一系列观测所反映的不过是在这个巨大、无限的宇宙网上的一部分宇宙中所发生的实在性，而每一个这样的宇宙中都有其他版本的你、我以及每一个生活在这个宇宙中的人；在这些人生活于其中的宇宙中，一定的观测还是会带来一定的结果。在这个宇宙中，你在看这些字，在另一个宇宙中，你在休闲上网，而在另一个宇宙中你正紧张地等待着你在百老汇舞台上的首次演出。看起来并不存在图5.1所勾勒的单独一个时空块，似乎存在着无限多个时空块，每一个都代表着事件的一种可能性。在多世界理论中，可能出现的结果并不只是一种可能。波函数不会塌缩。每一种可能的结果都会出现在平行宇宙的某一个中。

20世纪50年代大卫·玻姆（我们在第4章中讨论爱因斯坦波多斯基罗森时曾提到过这位物理学家）提出的第3种设想是一种完全不同的想法。[8]玻姆认为，粒子，比如说电子，就像经典物理学中的观念以及爱因斯坦所希望的那样，的确具有确定的位置和速度。但是，为了与不确定原理相一致，这些性质被隐藏起来；它们是第4章中提到的各种隐变量的鲜活例子。你不能同时测量它们。对于玻姆而言，不确定性代表的只是我们认知上的局限性，而非粒子本身的属性。他的方法并没有违背贝尔的结果，因为就像我们第4章结尾所讨论的那样，具有量子不确定性原理所禁戒的确定性质这件事并没有被排除掉；被排除掉的只是定域性，而玻姆理论并非定域性理论。[9]玻姆另辟蹊径，将粒子的波函数想成另一种单独的实在性元素，一种独立于粒子本身而存在的元素。就像玻尔的互补性哲学的说法：既不是粒子也不是波；根据玻姆的观点，既是粒子又是波。而且，玻姆提出，粒子的波函数与粒子本身相互作用——它“引导”或“推动”粒子——波函数在某种方式上决定了粒子下面的运动。这种观点与标准量子力学成功的预言完全一致；玻姆发现，波函数在某个位置的变化会立即推动一个遥远位置上的粒子，这个发现清楚地说明了玻姆理论的非定域性。举个例子来说，在双缝实验中，每个粒子穿过这条或那条缝隙，而其波函数则两条缝隙都要穿过并且发生了干涉。既然波函数会引导粒子的运动，那么我们就无须因为方程告诉我们粒子更有可能落在波函数较大的位置而不太可能落在波函数较小的位置而感到惊奇，这样就解释了图4.4中的数据。在玻姆的方法中，并不存在单独的波函数塌缩阶段，如果你测量粒子的位置，发现它在这儿，那么在测量之前的那一刻粒子肯定就在那里。

第4种想法，是由意大利物理学家詹卡洛·吉拉蒂、艾尔波特·里米尼、图里奥·韦伯提出来的，他们以一种巧妙的方式大胆地修改了薛定谔方程，同时却对单个粒子的波函数演化没有什么影响，只有将新的方程应用于“大”的日常生活中的物体时才会对量子演化产生戏剧性的影响。这个修正版本认为波函数本来就是不稳定的；这些人提出，即

使没有任何干预，每个波函数迟早也会按自己的节奏自动塌缩成峰状。吉拉蒂、里米尼和韦伯提出，对于单个粒子而言，波函数的塌缩会自发且随机发生，平均说来，每10亿年大约只发生1次。[10]塌缩发生的频率太小了，以至于不会使单个粒子的常规量子力学描述有什么改变，这非常好，因为量子力学以前所未有的精确性描述了微观世界。但对于实验者和他们的仪器这种由数以亿计的粒子组成的大物体而言，情况就不一样了；由于粒子数量极多，因而在极小的时间内，都至少有可能有一个组分粒子的波函数自发塌缩，从而使其波函数发生塌缩。就像吉拉蒂、里米尼、韦伯和其他人论证的那样，一个大物体中所有单个波函数的纠缠性质使得该种粒子的波函数的塌缩引起了量子的多米诺效应：所有组分粒子的波函数都发生了塌缩。由于这一切只发生在一眨眼的工夫，吉拉蒂、里米尼和韦伯所提出的修正版确保了大物体总会处于确定的状态：测量仪器上的指针总是指向一个确定的值；月亮总是在天空某个确定的位置；实验者的大脑中总有确定的体验；猫只能要么死了，要么活着，两者必居其一。

以上所述的每一种方法，以及一些我们在这里没有讨论的其他方法，都自有其支持者和反对者。“把波函数当做认知”这种方法否定波函数的实在性，仅把波函数视作我们所知的一切的说明符，从而巧妙解决波函数塌缩的问题。但是反对者会问，基本物理为什么非得与人类意识联系得如此紧密？如果我们没在观测这个世界，波函数是不是就永远不会塌缩？或者说，波函数这个概念是不是就不存在呢？在地球上的人类进化出意识以前，宇宙会不会完全是另外一个样子？如果观测者不是人类而是老鼠、蚂蚁、变形虫或者电脑之类，那又会有什么不同？其“认知”上的变化大到足以与波函数塌缩联系起来吗？[11]

与之相比，多世界诠释规避了整个波函数塌缩概念，因为在这种方法中波函数不会塌缩。但代价是存在无数个宇宙，而这是令很多反对者不能接受的事情。[12]玻姆的观点同样规避了波函数塌缩；但其反对者认为，如果同时赋予粒子和波以独立实在性，那这个理论未免不

太经济。而且，反对者们正确地指出了在玻姆的体系中，波函数对其所推动的粒子的影响速度比光还快。其支持者们则认为，前一种抱怨可算是主观性的，而后者又符合贝尔所证明的不可避免的非定域性，因此这两种批评意见都没什么说服力。然而，对玻姆可能不太公平的是，其方法从没有流行起来。[13]吉拉蒂里米尼韦伯的方法通过改变方程使其包含一种新的自发塌缩机制从而直接解决了波函数塌缩问题。但反对者们指出，还没有实验证据支持其对薛定谔方程的修改。

为寻求量子力学的形式主义与日常生活经验之间可靠而又完全清晰的联系所做的研究无疑会一直进行下去直到问题得以解决，现在还很难说到底哪种现有方法会最终得到大多数人的认可。要是物理学家们今天就投票，我认为不会有哪种方法获得压倒性的优势。不幸的是，实验数据帮不上什么忙。吉拉蒂里米尼韦伯的方法确实给出了在某些情况下不同于标准的阶段一阶段二量子力学的预言，但偏差非常之小以至于无法用今天的技术加以验证。其他3种方案的情况就更加糟糕了，因为它们更加明确地抗拒实验检验。它们都与标准方法一致，因此对可进行的观测和测量，都只能给出同样的预言。它们之间的区别只在于幕后发生的事情不同。也就是说，它们之间的区别只表现在用量子力学解释实在性的潜在性质时的不同。

即使量子测量问题还没有解决，在过去的几十年间，一种基本框架却一直在发展中，尽管还不完善，但却得到了广泛的支持，被认为很可能是可行的解决方案的一个组成部分。这就是所谓的退相干。

## 退相干和量子实在性

当你初次遇到量子力学的概率时，自然的反应会是它并不比掷硬币或轮盘赌中的概率更为奇妙。但当你了解量子干涉时，你会意识到概率是以一种更为基本的方式进入量子力学中的。在日常例子中，各种与概率有关的结果——正面与反面，红与黑，一个抽奖数字与另一个抽奖数字——都可以这样理解：最终一定会出现这种或那种结果，

而每一种结果都是一段独立而又确定的历史的最终产物。掷硬币时，有时旋转运动正好使得正面向上，有时又恰好是反面向上。每种结果50%的概率并不只与最终结果——正面还是反面——有关，还与导致每种结果的历史有关。你有一半的机会掷出正面向上的硬币，也有一半机会使硬币反面向上。这两种历史本身完全分离，各自独立。不同的硬币运动既不会彼此增强也不会彼此抵消。两种历史全都是独立的。

但在量子力学中，事情是不一样的。电子从双缝到探测器所走过的各种路径并不是分离的、孤立的历史。各种可能的历史混合起来产生可观测结果。有些路径会彼此增强，有些路径会彼此削弱。正是各种可能历史之间的量子干涉使得探测屏上出现明暗相间的图样。因此，量子物理概率概念与经典物理概率概念之间的区别在于，前者可归结为干涉效应，而后者则并非如此。

退相干性是一种普遍存在的现象，通过压低量子干涉——也就是说强烈地削弱量子概率和经典概率之间的核心差异，退相干架起了小小世界的量子物理和没那么小的世界的经典物理之间的桥梁。早在量子理论的早年岁月，人们就已经认识到了退相干的重要性，但其现代形式则可追溯到德国物理学家迪尔特·泽尔1970年的一篇开创性文章，[14]之后，包括德国的埃里克·乌斯，美国新墨西哥州洛斯阿拉莫斯国家实验室的沃切克·祖莱克在内的一些物理学家进一步发展了这一理论。

主要思想是这样的，当将薛定谔方程应用于简单的情况，比如通过有双缝的屏幕的单个独立光子，就会形成著名的干涉图样。但实验室中的实验有两个特别之处是真实世界所无法具有的。第一，我们在日常生活中所遇到的事物要比单个光子大得多，复杂得多。第二，我们在日常生活中遇到的事物并不孤立：它们总与我们及周围的环境相互联系。现在在你手中的这本书就与人类有接触，更一般性地说，这本书正持续不断地被光子和空气分子撞击。而且，由于书本身是由许

多分子和原子组成的，这些躁动不安的组分本身也会互相碰撞。同样的道理也适用于测量仪器上的指针、猫、人类的大脑，以及你在日常生活中碰到的每一件事物。在天体物理中，地球、月球、小行星以及其他行星不断地被来自太阳的光子撞击。甚至是漂浮在漆黑的太空中的一粒灰尘也不断受着宇宙大爆炸以来遍布于空间的低能微波光子的撞击。因此，为了理解量子力学怎样解释真实世界中的事物——而不仅是原始的实验室中的实验——我们应把薛定谔方程应用到更加复杂、更为麻烦的情况中去。

从本质上来看，这就是泽尔所强调的，而他本人的工作以及其后的许多其他人工作揭示了一些不寻常的事情。虽然光子和空气分子如此之小以至于对书、猫之类的大个物体不会产生什么实质性的影响，但它们会有别的作用。它们不断地“推动着”大物体的波函数，或者用物理术语讲，它们干扰着大个物体的干涉性：它们扰乱了波峰波谷的排列顺序。这一点很关键，因为波函数的有序性对于产生干涉效应是非常重要的（图4.2）。正如将标记装置添加到双缝实验后，由于扰乱了波函数，所以消除了干涉效应；环境中的成分持续不断地撞击物体也有消除干涉现象的可能性。反过来看，一旦量子干涉不再可能，量子力学所固有的概率性，从实际的角度看，就会像掷硬币或轮盘赌所固有的概率性一样。一旦环境的退相干性弄乱了波函数，量子概率的奇异性就会变成日常生活中所熟悉的概率。[15]这表明我们有可能解决量子测量之谜，而这将是大家期待见到的最激动人心的事。接下来我将首先要以最乐观的态度讲讲它，然后再强调还需要做哪些事。

假如一个孤立电子的波函数表明它有50%的概率在这儿，有50%的概率在那儿，则我们必须用量子力学发展成熟的奇异性质来诠释这些概率。由于两种情况皆可通过混合并生成干涉图样来展现自己，我们必须将两者视为同等真实。不那么严格地说，这就意味着电子处于两个位置。如果我们用非孤立的、日常大小的实验仪器来测量电子位置，将会发生什么呢？这个嘛，与电子不确定的位置相对应，测量仪

器上的指针也会有50%的概率指向这个值，50%的概率指向另一个值。但由于退相干性，指针不会指向两个值的混合值；由于退相干性，我们可以从通常的、传统的、日常的意义上来诠释这些概率。就像一枚硬币有50%的概率正面朝上，有50%的概率反面朝上，但不能确定是正面朝上或反面朝上，测量仪器的指针有50%的概率指向这个值，有50%的概率指向另一个值，但会明确地指向一个值或另一个值。

类似的论证也适用于其他更为复杂的非孤立对象。如果量子计算告诉我们，一只猫坐在密闭盒子里，有50%的概率死掉，有50%的概率活着——因为一个电子有50%的概率撞上盒子里的陷阱使猫吸进毒气而死亡，也有50%的概率幸运地避过陷阱——那么根据退相干性，猫不会处在既死亡又活着的荒唐的复合状态。虽然几十年来，人们一直在热情不减地讨论着这样一些问题，比如，猫处于既死亡又活着的状态究竟是什么意思？打开盒子的行为和观察猫的行为究竟是如何迫使其选择死或活这样的确定状态的？退相干性却告诉我们，早在你打开盒子之前很久，环境已经完成了无数次观测，并立刻将所有的神秘的量子概率转化为毫无神秘可言的经典对应。在你看猫之前很久，环境已经迫使猫处于一种唯一的确定的状况。退相干性迫使量子力学的许多古怪之处从大个物体中“漏网”，因为量子的古怪之处被来自环境中的无数粒子的碰撞一点一点除去了。

很难想象有更加令人满意的解决量子测量问题的方法。更为现实一点并且放弃忽略环境因素的简单假设——在该领域的早期发展阶段，简化处理对于取得进展十分重要——的话，我们将发现量子力学有一个内在的解决之道。人类的意识、实验者和人类的观测不再起特殊作用，因为它们（我们）只不过是像空气分子或光子一样的环境元素，这些东西在给定物理系统中可以相互作用。在观测对象的演化和做观测的实验者之间，也不再会有阶段一、阶段二的划分。每一样事物——被观测者和观测者——处于同等的地位。每一件事物——被观

测者和观测者——都可由薛定谔方程所决定的一模一样的量子力学定律掌控。测量行为不再特别，它只不过是与环境发生作用的一个特殊例子而已。

就这样吗？退相干性解决了量子力学测量问题吗？使波函数关闭其他所有可能实现的可能性而只保留其中一种的是退相干吗？有些人认为是这样。研究者们，如卡内基·梅隆的罗伯特·格里菲思，奥尔塞的罗兰德·奥内斯，圣达菲大学的诺贝尔奖获得者穆雷·盖尔曼和加州大学圣巴巴拉分校的吉姆·哈特尔，取得了巨大进展并声称他们已经将退相干发展成了可以解决测量问题的完整理论框架（被称为退相干历史）。其他人，比如我自己，对这个问题非常感兴趣但是还不完全相信。你看，退相干性的强大之处在于，它成功地破除了玻尔在大小物理系统之间设置的人工障碍，使每一样事物都可以被纳入同一套量子力学体系。这一进展非常重要，我想玻尔也会感到满意。虽然尚未解决的量子测量问题并未削弱物理学家们用实验数据验证理论计算的能力，但它却使玻尔和他的同事们一起制定了一套有着明显不妥当的性质的量子力学体系。许多人发现，这个体系需要一些诸如波函数塌缩或原属于经典物理领域的“大”系统的不准确概念之类的含糊词语，而这令人无法完全信服。某种程度上讲，对退相干性的思考使研究者们认识到那些含混不清的思想没必要存在了。

但是，我在上面的讲述中避开了一个关键问题：就算退相干性抑制了量子干涉，进而使量子概率像熟悉的经典物理对应一样，波函数中每一潜在结果仍有可能成真。因此我们仍然好奇一种结果是如何“胜出”的？而当胜出的可能性成真时，其他可能性又是如何“退散”的？掷硬币时，经典物理也会回答类似问题。如果你知道硬币旋转的准确方式，理论上讲，你可以预言它是正面落地还是反面落地。进一步思考发现，每一个结果都是由你最初忽略的细节所精确决定的。在量子物理中我们不能说类似的话。退相干性允许我们用类似于经典物理的方式来诠释量子概率，但却没有再为我们提供任何细节，

使我们知晓究竟是怎样从很多可能结果中挑出一种使其实际发生。

在精神实质上，有些物理学家和玻尔一样，认为寻求这样一种解释——用以说明单一确定的结果是如何出现的——是一种误导。这些物理学家认为，量子力学及其包括了退相干的升级版，是一种结构很严谨的理论，其预言可以解释实验室中测量装置的行为。在这种观点看来，这就是科学的目的。为究竟发生了什么寻求一种解释，努力理解一种特殊的结果是怎样出来的，追寻在一定程度上超出了探测器读数和电脑结果的实在性所暴露出来的是非理性的智力贪欲。

另外的许多人，包括我自己在内，持有一种不同的观点。解释数据的确是科学。但很多物理学家们相信，科学也应包含那些实验数据证实的理论，进而利用它们来获得对实在性本质的最大领悟。我坚定地相信，在寻求测量问题的完美解决方案的驱动下，我们会获得更为深刻的领悟。

因此，虽然人们普遍认为环境诱发的退相干性是跨越量子物理经典物理分界的关键，而且许多人也希望这些想法有朝一日能在量子物理与经典物理之间搭建一条完善且具有说服力的桥梁，但是大家都觉得这座桥梁还远没有建好。

## 量子力学和时间之箭

我们站在测量问题的哪里？测量问题对时间之箭又意味着什么？宽泛地讲，把我们的日常经验和量子实在性联系起来的方案有两大类。第一类（比如说，将波函数作为认知；多世界理论；退相干性），薛定谔方程是整个问题的根本；这类方案只是提供了不同的用以说明薛定谔方程对物理实在性意味着什么的方式。第二类（比如说，玻姆理论；吉拉蒂里米尼韦伯方法），薛定谔方程必须用其他方程（在玻姆理论中，所需的方程展现了波函数如何影响粒子）来加以补充或加以修正（在吉拉蒂里米尼韦伯方法中，需要包含一种新的、

明确的塌缩机制）。判定测量问题对时间之箭的影响的关键问题是，这些方案是否引进了时间方向上的基本不对称性。别忘了，薛定谔方程就像牛顿、麦克斯韦和爱因斯坦的方程一样，完全同等的对待时间上向前和向后的方向。它并没有为时间的演化指明方向。这些方案会改变这种局面吗？

在第一类方案中，薛定谔的体系完全没有被修改，因此会继续保持时间上的对称性。在第二类方案中，时间对称性可能存在也可能不存在，这得取决于细节问题。举个例子来说，在玻姆理论中，新提出的方程同等地对待过去和未来，因此不会引进非对称性。但是，吉拉蒂、里米尼和韦伯提出的塌缩机制中确实存在着时间之箭——“不塌缩的”波函数，从峰形变化到延展状的波函数，并不符合修改后的方程。因此，根据方案的不同，量子力学以及量子测量之谜的解决方案，可能会也可能不会平等地对待时间上的不同方向。下面让我们来考虑一下每种可能性的含意。

如果时间对称性得以保有（我怀疑是这样），上一章中的所有论证和结论不用改变多少就可以应用到量子领域中。有关时间之箭的讨论中的核心物理概念是经典物理中的时间反演对称性。虽然量子物理的基本语言和结构不同于经典物理——波函数代替了位置和速度；薛定谔方程代替了牛顿定律——但所有量子方程中的时间反演对称性保证了对待时间之箭的方式不会改变。只要我们用波函数的语言描述粒子，我们就可以像在经典物理中那样定义量子世界中的熵。而熵总是在增长——朝我们称之为未来和过去的方向上都会增长——这一结论将得以保有。

因而，我们还会遇到在第6章中曾遇到的谜题。如果我们将此刻对世界的观测视为不容置疑的客观实在，如果熵同时朝着过去和未来的方向增长，那么我们该怎样解释世界为何会是现在这个样子，又将如何解释它接下来会怎样演变？于是又会出现同样的两种可能性：或者我们所看到的一切只是由于统计上的侥幸——在一个大部分时间内

都处于完全无序状态的永恒宇宙中，你会认为这种侥幸时不时会来上一次——而成为现实；或者由于某种原因，宇宙在大爆炸之后的熵出奇的低，140亿年来事物缓慢演变，而且会朝着未来的方向继续这样走下去。就像在第6章中那样，为了避免陷入不信任记忆、记录和物理定律的困境，我们再次将目光聚焦于第二种选择——低熵的大爆炸——并且为事物怎样以及为什么会开端于这样一种特殊状态这一问题寻求一种解释。

从另一方面来讲，如果时间对称性不复存在——如果最终被接受的测量问题的解决方式表明，量子力学在处理未来和过去上存在着基本的不对称性——那么我们就能为时间之箭提供最直接的解释。比如说，测量问题的解决可能会告诉我们，鸡蛋之所以会破碎而不会聚集起来是因为，不同于我们用经典物理定律发现的那样，打碎的鸡蛋才是完整量子方程的解，而没打碎的鸡蛋则不是。逆向放映打碎鸡蛋的片子，会使我们看到不会在真实世界中发生的事情，这也就解释了为什么我们从未看到过这样的事。事实就是这样。

可能吧。虽然看起来这似乎为时间之箭提供了一种全然不同的解释，但事实上却并非那么不同。就像我们在第6章中强调过的，要想使《战争与和平》的页码越来越乱，则一开始页码必须有序排列；要想通过打碎鸡蛋的办法使其变得无序，一开始就得有一个有序的完好的鸡蛋；熵朝着未来的方向增加，是因为熵在过去很低，因此它有潜力变得更加混乱。然而，某定律对待过去和未来的方式不同，并不能说明该定律规定了过去的熵应该很低。该定律也可能意味着过去的熵应更高（或许熵会在未来和过去这两个方向上呈不对称增长），甚至有可能时间不对称定律根本不能说明过去怎样。后者正是吉拉蒂里米尼韦伯方案中的情况，而这个方案是目前市面上唯一具有实质性的时间不对称性的方案。一旦他们的塌缩机制生效，就没有办法将其撤销——不可能从塌缩的波函数开始，令其演化到之前延展的形式。波函数的细节已经在塌缩过程中遗失了——它变成了峰状——因此不可能

使事物“重返”其波函数塌缩之前的任何时刻的样子。

因此，即使时间不对称性定律部分地解释了事物为什么只能朝时间的一个方向演变而从不能反过来演变，它仍然同时间对称性定律一样，需要关键性的补充：解释清楚为什么在遥远的过去熵会很低。当然，这是目前为止提出过的对量子力学所做的真正的具有时间不对称性的修改。因此，除非未来的发现弄清楚了有关时间对称性或不对称性的问题，而我认为这都不太可能——那么量子测量问题的时间不对称性解，将会保证在朝着过去的方向上熵会减少——否则的话，我们试图解释时间之箭的努力将又一次把我们带回到宇宙起源的问题上，这个问题将是本书的下一部分所要讨论的内容。

在这几章中我们已清楚地看到，宇宙学的思考盘绕在有关空间、时间和物质的问题的神秘核心地带。因此，在用现代宇宙学思想探寻时间之箭的奥秘的旅途中，我们不要走马观花，而要漫步于宇宙历史中细细探究。

# 第3部分
# 时空与宇宙学

## 第8章　雪花与时空

对称性与宇宙的演化

理查德·费曼曾经说过，如果让他选择一句话来概括现代科学中最重要的发现，他会选“世界是由原子组成的”这句话。一旦我们认识到我们关于宇宙的诸多知识都是建立在原子的性质和相互作用理论的基础之上——无论是解释星星为什么发光，天空为什么是蓝色的；还是解释你的手为什么能感觉到这本书，眼为什么能看见上面的字，我们都需要用到原子的知识——我们就会明白费曼的选择是多么的明智。许多当代最著名的科学家认为，如果有再选一句话的机会，那么所选的将是“对称性是宇宙规律的基础”。过去的数百年间，科学领域的巨变难以记数，但其中最有生命力的发现具有某种通性：这些发现的着眼点都是那种在多种操作下具有不变性的大自然性质。这种不变的特性就是物理学家们所说的对称性。在很多重要的科学进展中，对称性都在扮演着非常重要的角色。我们有足够的理由相信，躲藏在其神秘面纱之后的对称性，将以耀眼的光辉照亮有待发现真理的黑暗角落。

事实上，我们将会看到，宇宙的历史在某种程度上可说是对称性的历史。宇宙演化中最关键的时刻就是平衡与秩序被突然打破的时刻，在这样的时刻，宇宙的性质会突然变得不同于之前。根据现代理论，宇宙在其诞生之初的一段时间内经历过数次巨变；我们今日所见到的一切事物都是极早期高度对称的宇宙所残留下来的遗迹。而按照

更为深刻的理解，对称性根本就是宇宙演化的关键。时间本身就与对称性密切相关。我们将在后面看到，从实践中得来的作为变化的量度的时间概念，以及宇宙演化过程中某一段特定时期的存在——关于这样的时期，我们可以谈论一些诸如“宇宙作为一个整体的年龄和演化”这样的内容——都与对称性的有关方面密切相关。当科学家们研究宇宙的演化，追本溯源探求空间和时间真正的性质时，对称性已向我们证明它就是最佳向导，只有它才能帮助我们洞悉那些完全无法触及的真相，找到答案。

## 对称性与物理定律

对称性无处不在。我们玩台球的时候每次都要击打的白色主球，拿起它，随便怎样转一下它，绕哪个轴都行，它看起来还是原来的那个样子。让一个没有花纹的圆盘子绕着它的中心转，它看起来在转动中没有任何改变。轻轻地拿起一片刚落下的雪花，把它的每个角转到相邻角的位置，你会发现很难看出这片雪花经过了转动。让一个字母“A”绕着穿过其顶端的垂直轴翻转一下，你将得到一个看起来一模一样的“A”。

这些例子很清楚地告诉我们，一个物体所具有的对称性指的是一种操作，不管这种操作是真实的还是想象的，只要在这种操作下，该物体看起来没有发生任何变化，我们就可以将这种操作称为该物体所具有的对称性。对于一个物体来说，能令其保持不变的操作种类越多，它所具有的对称性就越多。完美的球体具有高度的对称性，因为任何一种转动，不论是绕上下贯通的轴，还是绕左右贯通的轴，又或者是绕前后贯通的轴，只要其轴经过球体中心，该转动都无法使球有任何变化。立方体的对称性则要少一些，因为只有绕垂直于立方体表面的中心轴（每根这样的轴同时垂直于两个面）旋转90度才能保持立方体不变。那么当然，一旦有人用其他方式旋转了立方体，比如按图8.1（c）中的那种方式，你仍然可以一下子认出那个立方体，但你也会同时发现有人碰过它了。而对称则像最老练的小偷，他们什么证据

都不会留下。

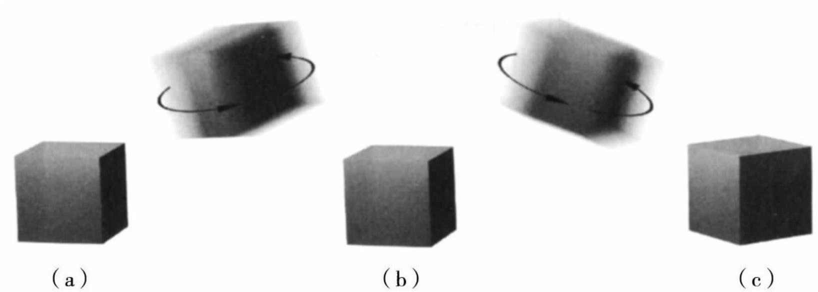



图8.1　图（a）中的立方体绕着其中一面的轴旋转90度或其整数倍，立方体保持不变，如图（b）所示；但旋转的度数若不是90度或其整数倍，立方体的旋转就能看得出来，如图（C）所示。

所有的这一切都是有关物体在空间中的对称性的例子。已知物理学定律中所暗含的对称性与这些例子密切相关，只不过我们要以一种更为抽象的方式提出这个问题：施加于你或环境上的哪一种操作——不论其为真实的还是想象的——会对那些用以解释你所观测到的物理现象的定律没有任何影响？值得注意的是，所谓的对称性操作，并不是要求保持你的观测不变。我们真正关心的是，支配这些观测的定律本身在对称性操作下是否不变；也就是说，用以解释你做对称性操作之前的观测的定律，同用以解释你做对称性操作之后的观测的定律是否完全一样。因为这是我们所讨论的中心思想，所以让我们花点时间看看某些例子。

让我们将你假想为一名体操运动员，过去4年间你一直在康涅狄格运动中心为准备奥运会而进行训练。经过日复一日的重复训练，你已经可以轻松地完成你的体操套路中的每一组动作——你很清楚在平衡木上需要用多大的劲才能完成挺身前空翻，在地板上跳多高才能完成一个直体后空翻转体720度，在双杠上把身体摆多快才能完成一次完美的空翻两周下。看起来，你的身体对牛顿定律有本能的感觉，因为正是这些定律支配着你身体的运动。现在，在纽约举办的奥运会真

正开场了，你要在现场观众面前表演你的套路，你当然希望牛顿定律保持不变，因为你想展现出来的是与练习时完全一样的套路。而如我们所知，牛顿定律的确可以满足你的这种期待。它并不会随位置的改变而变化。纽约与康涅狄格不会有两套牛顿定律。我们相信，不论你在哪，牛顿定律都会是一个样子。即使你换个地方，支配你身体运动的定律也不会受到任何影响，就像那颗台球在旋转时看不出表面有任何变化一样。

这一对称性被称为平移对称性或平移不变性。平移对称性不仅在牛顿定律中成立。在用来描述电磁相互作用的麦克斯韦定律中，爱因斯坦的狭义相对论和广义相对论中，量子力学中，以及现代物理学形形色色的理论中，平移对称性都是成立的。

这里需要注意一点。你的观测与体验会因为你所处位置的变化而不同。如果你在月球上完成你的体操套路，你就会发现，当你的双腿使同样的力气向上跳时，身体弹起的轨迹却与在地球上时完全不同。当然，我们懂得其中的差异，知道这正是物理定律导致的。月球比地球轻很多，因而能够产生的引力也就小很多，从而使得你向上蹦起的轨迹与在地球上时完全不同。而这一事实——引力的大小取决于质量的大小——正是牛顿引力定律（以及更加精确的爱因斯坦广义相对论）的一个组成部分。你在地球上及月球上感受的不同并不代表引力定律随地点的变化而变化。相反，真正体现出来的只是环境的变化。所以，当我们说已知的物理定律本身不会因为康涅狄格和纽约——我们甚至可以把月球也加进去——的区别而有变化时，必须同时记得定律也依赖于环境的差异。总而言之，我们要记住的关键性结论是，用以解释自然现象的物理定律的框架绝不会随着位置的改变而发生变化。地理上的改变不会逼迫物理学家回到黑板前重新推导理论。

物理学定律不一定非得这样。我们也可以臆想出一个新宇宙，其中的物理定律就像地方及国家政府一样随时随处变化；我们在地球上所熟知的物理定律完全不能帮助我们了解月球、仙女座星系、蟹状星

云或是宇宙中其他位置的物理定律。事实上，我们并不真的那么确定在我们这里起作用的物理定律是否真的也在宇宙其他角落有效。但是我们的确知道，要是宇宙中某处的物理定律不同于我们所想，那它必须在那里找个出口把这种差异消化干净。因为越来越精确的天文学观测事实已向我们提供了足够可信的证据，证明在整个宇宙空间中，或者更准确地说至少在我们目前所能看到的宇宙空间中，物理定律是一致的。这一点更加突出了对称性的神奇威力。虽然我们只能在地球及其附近活动，但空间平移不变性的存在，却使我们能够足不出户就洞悉整个宇宙的基本定律，因为在我们这里发现的物理定律同时也是整个宇宙的定律。

转动对称性或转动不变性与平移不变性本是近亲。这一对称性基于这样一种理念：不同的空间方向有相同的地位。在地球上的观测并不能使我们得出这样的结论。我们抬起头看到的景象与低下头时看到的完全不同。但是同样的，这也仅仅是环境的细微差别，而非其背后的物理定律本身特性的不同。如果你离开地球，漂浮在外太空，远离任何星星、星系或是其他重天体，转动对称性就会凸显出来；在黑洞洞的宇宙空间中，你找不出一个特殊的方向，四周全是一样的。要是你打算建造一个用以探索物质或力的性质的外太空实验室，那么你根本不必花心思在它的朝向问题上，因为基本定律根本不会被实验室朝向影响。要是哪天晚上某个捣蛋鬼打算改变一下实验室回转仪的设置，使其按一定角度绕某个方向的轴旋转，你也用不着担心这会对你探索物理定律的实验有什么影响。人类至今所完成的所有实验都可以证明这一预期。所以我们相信，掌控着你所进行的实验以及用以解释你所得到的实验结果的物理定律并不在乎你在哪里——平移对称性，以及你面朝哪里——转动对称性。[1]

正如我们在第3章中所讨论过的，伽利略及其他物理学家深刻认识到物理学定律还应当遵守另一种对称性。如果你的外层空间实验室以匀速运动——不论你以每小时5000米的速度运动还是每小时10万

千米的速度运动——那么这种运动绝不会对用以解释你的实验观测的物理定律有任何影响，因为你与相对你静止的人不会有不同的观测结论。我们已经看到，爱因斯坦用一种完全不可预见的方式扩充了这一对称性，他提出无论相对于哪个观测者，光速都有确定的大小，绝不会因你或者光源的速度改变而改变。毫无疑问，这相当令人吃惊。因为一般情况下，我们认为一个物体的速度应该取决于其相对于外界环境的速度，观测到的速度依赖于观测者本身的速度。但是爱因斯坦从牛顿理论的缺陷顺藤摸瓜，发现了光的对称性，将光速提升到了不可侵犯的大自然定律层次，宣称其并不受运动的影响，正如白色的台球不会因旋转而改变一样。

爱因斯坦的下一项重大科学贡献——广义相对论，正是沿着这样的方向朝着具有更大对称性的理论继续前进。正如你可以将狭义相对论想成是在相对于彼此匀速运动的观测者之间建立对称性，你也可以将广义相对论想成是前进了一步的狭义相对论，它在相对于彼此加速运动的点之间建立对称性。这一点非常特别，因为我们强调过：你不能感受到匀速运动，但是你可以感受到加速运动。因而，描述你的观测的物理学定律看起来应该会因为你的加速而变得有所不同，以便解释你所感受到的额外的那部分力。而这正是牛顿理论的情况。我们在一年级物理课程中学习的牛顿理论，在加速情况下必须有所修改。而我们在第3章中讨论过的等价原理，则使爱因斯坦认识到，你无法分辨出在加速过程中所感受到的力同处于相应大小的引力场（加速度越大则相应的引力场也应当越强）时所感受到的力之间的差别。爱因斯坦凭借其精深的洞察力认识到，一旦将合适的引力场添加到外在的物理条件中，在你加速时物理定律就不会发生变化。在广义相对论的框架下，所有的观测者，即使那些做任意大小变速运动的观测者，都具有平等的地位——他们彼此完全对称——因为每一个观测者都可以宣称自己处于静止状态，只要他们将各自所感受到的力算作不同的引力场的效应。这样一来，相对于彼此加速运动的观测者的观测事实之间存在差异就毫不为奇了，而且也不能再被算作是自然定律改变的证据

了，这就如同你在地球和月球上分别完成你的体操套路感受不同不能作为自然定律变化的证据一样。[2]

上面的这些例子使我们能够理解为什么很多人（我猜费曼也会同意）会认为在我们最深刻的科学认知排名中，自然定律背后的大量对称性可获得仅次于原子假说的亚军了。不过故事还远未结束。过去的几十年间，物理学家们将对称原理的地位大大提升到我们的科学探索之梯的最高一级。如果有人提出一条新的自然定律，我们就会很自然地问出：为什么要有这条定律？为什么要有狭义相对论？为什么要有广义相对论？为什么要有麦克斯韦电磁理论？为什么要有关于强相互作用和弱相互作用的杨米尔斯场论（我们稍后再来谈这个理论）？回答这些问题时很重要的一点就是要知道这些理论的预言可以被精确的实验反复验证。这一点对于建立物理学家对这些理论的信心非常重要。但是除此之外，我们还得知道有一些其他的重要原因。

物理学家们之所以相信这些理论在正确的轨道上还有另外重要的理由，虽然不好形容，但是我们可以说是物理学家们感觉这些理论是正确的，而对称性的思想则对他们的这种感觉至关重要。正是因为没有理由认为宇宙中存在某一与其他位置相比独一无二的位置，所以物理学家们对平移对称性广泛地存在于自然定律中有信心。正是因为没有理由认为宇宙中存在某种与其他匀速运动相比独一无二的匀速运动，所以物理学家们有足够的信心将狭义相对论——在所有匀速运动的观测者之间建立对称性而得到的理论——视作自然定律的重要部分。更进一步，正是因为没有理由认为任何一个观测点——不管其加速与否——会不如其他观测点，所以物理学家们有足够的信心将广义相对论——能将这种对称性纳入囊中的最简单理论——视作掌控一切自然现象的基本真理。另外，我们即将看到，关于除引力之外的3种力——电磁力、弱核力与强核力——的理论，正是建立在另外一些更加抽象但却同样引人注目的对称原理的基础之上。所以，自然界中的对称性并非是自然定律的结果。按照现代观点，对称性是自然定律的

基础。

对称性与时间

对称性的思想不仅对于与自然界中的力有关的物理规律非常重要，对于时间本身也非常重要。没人能够给予时间一个明确的基本层面上的定义。但毫无疑问的是，时间在宇宙组成中的部分角色为变化的记录者。事物逐渐变化不同以往，使我们注意到时间的流逝。手表上的指针指向不同的数字，太阳在天空中的位置发生变化，你复印的《战争与和平》的页码因为没有装订而越翻越乱，可乐瓶中出来的二氧化碳四处弥漫——所有的这一切都表明事物发生了变化，而时间的作用正在于它可以帮助我们注意到这些变化。按照约翰·惠勒的说法，时间就是大自然用以保证所有的一切——所有的变化——不至于一股脑儿发生的巧妙方法。

时间的存在取决于一种特殊对称性的缺失：对我们来说，即使定义与我们的日常感知类似的时时刻刻的概念，也需要宇宙中的万物必须时时刻刻有所改变。如果在今日之世界与过去之世界之间存在一种完美的对称性，如果时间的改变就像旋转白色台球一样不会带来任何变化，那么我们所感知到的时间实际上就并不存在。[3]这并不是说图5.1中逐步展示的那种时空膨胀并不存在，它还是能够膨胀。但是因为时间轴上的一切都完全一致，所以宇宙的演化或改变这种说法就没意义了。时间只是这一实在性的舞台中的一个抽象概念——时空连续统的第四维——否则它就不可辨别。

然而，即使时间的存在与某种特别的对称性的缺失联系在一起，其在宇宙尺度上的应用则要求宇宙必须严格遵守另一种不同的对称性。其中的思想非常的简单并且可以回答你在阅读第3章时可能遇到的一个问题。既然相对论告诉我们时间的流逝快慢取决于你运动的速度以及你所处的引力场的强度，那么我们不禁要问：天文学家和物理学家谈起整个宇宙起始于某一特定时刻——今天的天文学家和物理学

家认为这个时刻差不多是140亿年前——时又是什么意思呢？这140亿年又是相对于谁来说的呢？哪一台钟给出的140亿年？遥远星系上的智慧生命也会得出宇宙的寿命是140亿年的结论吗？而要是这样的话，又是什么保证了他们的钟和我们的钟同步校对过呢？这些问题的答案都取决于对称性——空间中的对称性。

如果你的眼睛可以看见波长远远超过红光或橙光的波长的光的话，你就不仅会在按下启动按钮时看到微波炉内部突然放射微波开始烘烤的景象，还将在我们这些普通人眼中漆黑一片的夜空中看到虽然暗淡但却几乎均匀的红光。40多年前，科学家们发现宇宙中弥漫着微波——波长很长的光——辐射，这种微波辐射正是大爆炸刚刚结束时极度高热环境残留至今的冷却遗迹。[4]宇宙微波背景辐射完全无害。早期宇宙处于难以想象的高热状态，但随着宇宙的演化与膨胀，辐射被稳定的稀释，慢慢冷却了。今天，微波辐射的温度只比绝对零度高2.7度。它所能搞出的最大恶作剧就是无线电视在信号不好以及调到一个没有节目的频道时所出现的雪花点。

但是这一微弱的静电噪声之于天文学家却如暴龙骨之于古生物学者：一扇通往较早时期的窗口对于重构遥远过去的一切极端的重要。通过过去10年的卫星探测，人们发现了微波辐射的一个重要性质：微波辐射的分布极其均匀。不同天空区域的微波辐射之间的差异低于千分之一。要是在地球上，这样的小的差异将使天气预报毫无意义。因为如果雅加达的气温是85华氏度的话，你立即就可以知道阿德莱德、上海、克利夫兰、安克雷奇[38]或其他任何一个城市的温度会在84.999华氏度～85.001华氏度。而在宇宙尺度上则完全不是这样，辐射温度的均匀性相当重要，之所以这么说有两个重要原因。

首先，辐射温度的均匀性提供了观测证据，证明宇宙在其早期并非由巨大的、高熵的物质团——比如黑洞之类——占据，因为这样参差不齐的物质环境只能留下同样参差不齐的辐射烙印。事实则相反，辐射温度的均匀性证明年轻的宇宙各向同性；而且，正如我们在第6

章中所看到的那样，与引力有关时——比如早期质密宇宙时引力起的作用——各向同性意味着低熵。这无疑是件好事，因为我们对时间之箭的讨论要求宇宙在其开天辟地时低熵。我们在本书的这一部分的目标之一就是尽我们所能解释这一观测事实——我们想要搞清楚早期宇宙的各向同性、低熵的这种非常不可能的状态是怎样出现的。这会使我们在时间之箭的探源之路上迈出一大步。

第二点，虽然宇宙演化自大爆炸，但平均下来，整个宇宙各处的演化应当彼此类似。既然我们这里的温度与涡旋星系、后发座星系团或者宇宙中的任意一处的温度都相同，那么太空中每一个地方的物理条件自大爆炸后一定按照相同的方式演化。这一推断非常重要，但是需要正确解释。仰望夜空，一眼看去我们会觉得天空并非一成不变：各种不同种类的行星与恒星闪耀天际。问题的关键在于，当我们对整个宇宙展开分析的时候，我们采用的是宏观视野，这些“小”尺度上的不同完全可以平均掉，在大尺度上宇宙的确是均匀的。我们只需想想简单的一杯水，在分子层次上，水是杂乱无章的：这里有个$H_2O$分子，那里又什么都没有，而另一边又有一个$H_2O$分子，等等。但是，如果我们对小尺度的水分子做平均，只考虑日常生活水平上这种肉眼可见的“大”尺度上的水，我们就会发现水是清澈均匀的。我们仰望星空时看到的不均匀正类似于水在分子水平上的不均匀性。但也正如用肉眼看那杯水一样，当我们在足够大的尺度上——以亿光年计数的尺度上——研究宇宙的时候，宇宙就具有高度的各向同性。因而辐射的均匀性就既是物理定律又是整个宇宙外在物理条件的均匀性的活化石。

这一结论意义重大，因为有了它我们就可以定义一个可用于整个宇宙的时间概念了。如果我们将变化的量度当成是时间流逝的一个有效定义，而整个空间的物理条件的均匀性就是贯穿整个宇宙的变化的均匀性的证据，那么我们就可以知道，时间流逝也具有均匀性。正如地球地理结构上的均匀性使得美洲的地理学家、非洲的地理学家以及

亚洲的地理学家彼此认同地球的历史与年龄，贯穿整个空间的宇宙演化的均匀性也会使银河系的物理学家、室女座星系的物理学家以及蝌蚪星系的物理学家得到一个大家都认同的宇宙历史与年龄。毫无疑问，宇宙演化上的各向同性意味着我们这里的钟与室女座星系的钟以及蝌蚪星系的钟，在平均的意义下，都取决于几乎类同的物理条件，所以会按差不多相同的方式计算时间。因而，空间上的各向同性保证了宇宙的同步性。

尽管我略掉了一些重要的细节（比如在下一节中将讨论到的空间的膨胀），这一段的讨论还是突出了问题的核心：时间在对称性下的尴尬处境。如果宇宙有短暂的对称性——如果其处于完全不变的状态——那甚至连定义时间都变得很难。另一方面，如果宇宙没有空间上的对称性——比方说，如果背景辐射完全是杂乱的，不同区域的温度有巨大的差别——宇宙学意义上的时间也就失去含义了。不同位置的钟表按照不同的快慢摆动，如果你要问一下宇宙30亿岁时是什么样子，那答案就将取决于你究竟是按照谁的钟来谈那流逝的30亿年。那将非常复杂。幸运的是，我们的宇宙既没有那么多对称性导致时间失去意义，也并非一点对称性没有使得我们无法避开那种复杂性，令我们无法探讨宇宙总的年龄以及随时间的总体演化。

那么现在让我们将目光转向演化，来一起思考宇宙的历史。

将结构放大再思考

宇宙的历史这几个字听起来像个大题目，但是这个大题目的纲要却极其简单，并且在很大程度上依赖于一个重要的事实：宇宙正在膨胀。既然这就是解读宇宙历史的核心要素，而且是人类最伟大的发现之一，那还是让我们来了解一下我们究竟是怎样认清它的。

1929年，埃德温·哈勃利用位于加利福尼亚州帕萨迪纳的威尔森山天文台100英寸望远镜，发现他所能探测到的几十个星系都在离他

远去。[5]事实上，哈勃发现，如果一个星系离他越远，则远去的速度就越快。为了对尺度有一个感性的认识，让我们看看哈勃原始观测方式的升级版（哈勃太空望远镜，研究对象为几千个星系）给出的数据：距离我们1亿光年远的星系以每小时550万千米的速度离我们远去；距离我们2亿光年的星系的移动速度也变成了2倍，即以每小时1100万千米的速度离我们远去；距离我们3亿光年的星系的移动速度就变成了3倍，即以每小时1650万千米的速度离我们远去，如此等等。哈勃的发现之所以令人震惊是因为按照当时主流的科学和哲学观念，最大尺度上的宇宙是静止的、永恒的、固定不变的。但是哈勃那一记重拳粉碎了这一观念。而在理论与实验的美妙结合中，爱因斯坦的广义相对论可以为哈勃的发现提供一个优美的解释。

事实上，你可能不会认为得出一个解释会异常困难。毕竟，要是你路过一个工厂，看到各种各样的材料猛烈又凌乱地四散飞了出来，你将不难猜出发生了一场爆炸。如果你再沿着金属和混凝土的碎块往回搜索，你会发现所有的碎片都聚敛于一个位置，而那很可能就是爆炸现场。按照相同的推理，既然从地球上看——哈勃与其后的实验都证明过的——所有的星系都在远离，你可能就会得出我们所在的位置正是远古时期大爆炸发生的位置，而各种恒星和星系就是那场大爆炸后均匀的四散飞出的碎片。此类理论的问题在于，其中必有一个特别的位置——我们所在的位置——是宇宙诞生的独一无二之地。要真的是那样的话，这个理论就必须承受一个根深蒂固的不对称性：远离爆炸核心的区域——也就是离我们很远的地方——与我们这里将会相当不同。因为天文学上的实验数据并没有为这种不对称性提供证据，而我们也高度怀疑这种以人类为中心的解释带有前哥白尼时代的气息，这就要求对哈勃的发现给出一个更加复杂的解释，在这个新的解释中，我们所在的位置不应当在宇宙中居于某种特殊地位。

广义相对论给出了一个这样的解释。利用广义相对论，爱因斯坦发现空间和时间是可变的，而不是固定的；是有弹性的，而不是刚性

的。他给出的方程，可以准确地告诉我们空间和时间如何随着物质和能量的存在而变化。20世纪20年代，俄罗斯数学家与气象学家亚历山大·弗里德曼和比利时牧师与天文学家乔治·勒梅特在将爱因斯坦理论应用于整个宇宙时各自独立地分析了爱因斯坦方程，并且得到了令人吃惊的结果。正如地球引力的存在使得接球手头上的棒球要么继续往上飞要么往下掉，而不会停在空中（除了到达最高点的那一瞬间），弗里德曼和勒梅特也认识到弥漫于整个宇宙空间的物质和辐射所具有的引力会使空间的结构要么拉伸要么压缩，就是不能保持固定不变的大小。事实上，这里用棒球做的比喻是极少数既抓住了物理本质又说清了数学内容的比喻。这是因为，掌控棒球离地面高度的方程同掌控宇宙大小的爱因斯坦方程非常类似。[6]

广义相对论中空间概念的灵活性为哈勃的发现提供了一种影响深远的解释方法。广义相对论没有按照工厂爆炸的宇宙学版本来解释星系的扩张运动，而是提出了空间本身已经膨胀了数十亿年。随着其自身的膨胀，空间也将星系拉离彼此，就像将生面团烤成松饼的过程中上面的芝麻点四散分离一样。因而，向外运动的起源并非是空间内部的爆炸导致的，而是空间自身持续不断地向外膨胀导致的。

为了更深刻地体会这一关键思想，我们再来想一下物理学家们常用来说明宇宙膨胀的非常有效的气球模型（这一类比的源头可追溯至一幅有趣的卡通画，可参见后面的注释。[7]这幅卡通画最早出现在一份1930年的荷兰报纸上有关威廉·德·西特的访问内容后，这位科学家在宇宙学领域作出了奠基性贡献）。这一类比将我们的三维空间同较易形象化的气球两维表面联系起来了，如图8.2（a）所示，其中的气球正被越吹越大。等间距粘在气球表面的硬币代表的就是星系。注意随着气球被吹起，硬币纷纷远离彼此，这个例子很形象地说明了膨胀的空间如何驱使星系远离彼此。

这一模型的一个重要性质在于，硬币之间是完全对称的，因为从每一枚硬币上的林肯像的视角上，看到的都是一模一样的景象。为了

更形象化一点，可以在脑海中将你自己缩小，躺到其中的一枚硬币上，然后看看气球表面上的所有方向（别忘了我们在这里将气球的表面类比成整个宇宙空间，所以如果你看的是除气球表面外的其他方向就没意义了）。你会看到些什么？你将看到气球上所有的硬币都会随着气球的膨胀而离你远去。换个硬币再试试你又看到了什么？对称性使你每次都只能看到相同的事情：

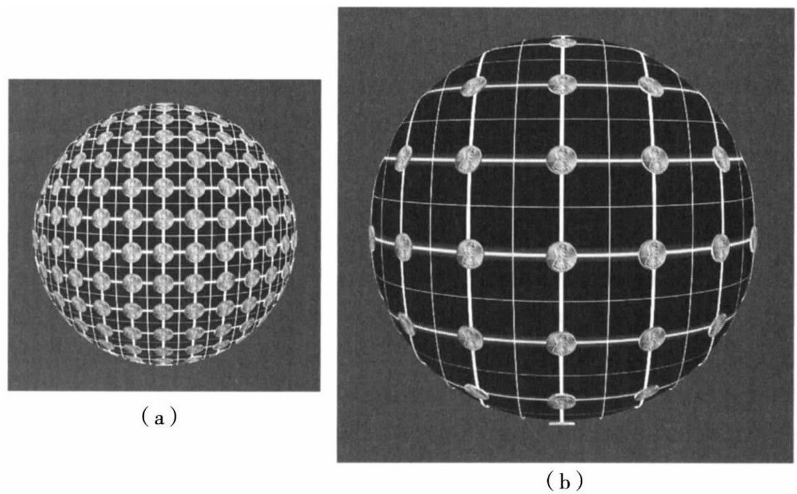


图8.2　（a）：我们把大量的1美分等间距的粘到球面上。如果每个林肯（译注：1美分上的林肯像）都观察其他林肯，那么他们会发现他们看到的景象完全一样。我们对宇宙的认识就是这样，从宇宙中的任何一个星系看去，所看到的景象与其他星系看到的景象平均说来是一样的。（b）：如果球面扩张，每个硬币之间的距离就会拉大。两个硬币如果在图8.2（a）中的距离越大，它们在图8.2（b）的扩张中拉开的距离就会越大。而我们对宇宙的观测也是如此，被观测点距离观测点越远，它离开观测点的速度也就越快。以上的讨论中我们并未假定有任何一个特殊的硬币存在，而我们的宇宙中也没有任何星系如此特殊以至于我们可以将它选为宇宙的中心。

所有方向的硬币都离你远去。这一切实的图像很好地符合了我们的信念（越来越多的精确天文学数据都在支持这样的信念）：宇宙中1000多个星系中的任何一处的观测者，当他们在强大的望远镜的帮助下凝望夜空的时候，在平均意义下，他们看到的图像与我们所看到的

会非常类似——周围的星系朝着所有的方向离我们远去。

所以，如果向外的运动起源于空间自身的膨胀，那么就不会像在一个固定的、先前即已存在的空间中的工厂爆炸事件那样需要一个特殊的作为向外运动的中心的点——没有特别的硬币，也没有特别的星系。每一个点——每一枚硬币，每一个星系——和其他的点具有完全等同的地位。任何一个位置的视野看起来都像是在爆炸的中心：每一个林肯都会看到其他的林肯远去；所有的观测者，无论在哪个星系，都会像我们一样，看到其他的星系远去。但因为对所有的位置都是如此，所以不会存在特殊的或者说独一无二的位置，不会有那个作为所有的向外运动起源地的中心。

此外，这一解释并非仅能用空间各向同性的方式定性地说明星系的向外运动，还能定量的符合哈勃的观测数据及其后更加精确的实验观测所给出的数据。如图8.2（b）所示，如果气球在一定时间间隔内向外膨胀，比方说大小增加了2倍，所有的空间间距也将变为2倍：之前相距1英尺的硬币现在就会相距2英尺，之前相距2英寸的硬币现在就会相距4英寸，之前相距3英寸的硬币现在就会相距6英寸，如此等等。因而，在任一给定时间间隔内，两枚硬币间距的增加正比于其初始间距。而因为给定时间间隔内间距的增加意味着速度的增加，所以离得越远的硬币彼此远离的速度就会越快。本质上，两枚硬币之间的距离越远，两者之间的气球面积就越大，所以气球膨胀时它们被推离的速度也就越快。将这一推导过程准确地应用于空间及其含有的星系的膨胀过程，我们就可以解释哈勃的实验观测了。两个星系的间距越大，则其间的空间就越大，因而空间膨胀时这两个星系被推离得也就越快。

广义相对论将观测到的星系运动归结为空间的膨胀，从而不仅提供了一个将空间中的不同位置平等处理的解释，还一下子说明了所有的哈勃实验数据。就这样，人们轻松地走到盒子外（这里的“盒子”指的是空间），用精准的数据以及奇妙的对称性来解释实验观测，这样

的阐释正是那种物理学家们会因其太过优美而不相信其可能出错的理论。空间结构正在膨胀这一猜想本质上符合全部观测。

## 膨胀宇宙中的时间

现在我们来看一个气球模型的变种。通过这个模型，我们可以更加准确地理解究竟怎样从空间——即使这里的空间指的是膨胀中的空间——的对称性来获得一个可以普遍应用于整个宇宙的时间概念。如图8.3所示，我们将图8.2中的硬币全部换成完全一样的钟表。根据相对论我们可以知道，如果这些完全一样的钟表所处的物理环境不同——处于不同的运动中或不同的引力场中——则它们所显示的时间变化快慢也将有所不同。但是简单地思考后我们即可知道，钟表之间将保留全部对称性，就像膨胀气球上的所有林肯那样。要是所有这些相同的钟表所处的物理条件一样，则将按照完全一样的快慢运转，记录的也将是完全一样的时间变化。与之类似，要是在一个膨胀的宇宙中所有的星系之间具有高度的对称性，那么随不同星系运动的钟表必将按照同样的快慢运转并且将记录下同样的时间变化。要不还能怎样呢？每一个钟表都等同于其他，平均说来，所有的钟表所处的物理条件几乎一样。这一点再次展现了对称性的强大。无须任何计算或者细致的分析，我们就可以认识到：物理环境的均匀性——通过微波背景辐射的均匀性以及整个空间中星系分布的均匀性体现出来[8]——使我们得到了时间的均匀性。

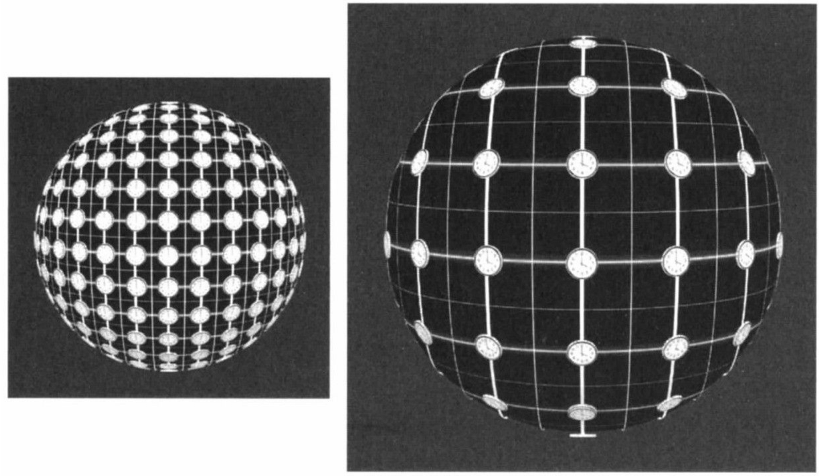


图8.3　随星系运动的钟表——其运动只取决于空间的扩张——可以当成是宇宙的钟表，虽然它们彼此相隔，但却可以保持同步，这是因为这些钟表随空间而动，而不是穿越空间而动。

尽管这一段的逻辑推理非常直接，其结论却令人困惑。既然所有的星系都随着空间的膨胀而快速远去，那么随着不同星系运动的钟表也将彼此远去。而且，它们相对于彼此远离的速度由于两两之间距离的不同而各不相同。那么这样的运动会不会就像爱因斯坦的狭义相对论告诉我们的那样，使这些钟表失去同步性呢？出于多方面的考虑，我们可以说答案并非如此。下面我们用一种特别有用的方式来想清楚整个问题。

回忆一下第3章中说过的，爱因斯坦发现按照不同路径穿过空间的钟表其指针快慢不尽相同（这是因为向着不同方向运动的钟表需要将不同时间长短的运动挪用为空间中的运动，还记得用滑板上的巴特打的那个比方吗？小家伙必须通过运动才能从朝北转向朝东）。但是我们现在所讨论的钟表并不穿越空间运动。就像粘在气球上的硬币那样只是随着气球表面的膨胀而相对于彼此远去，每个处于宇宙中不同位置的星系在很大程度上也只是随着整个空间的膨胀而相对于其他星系远去。而这就意味着，相对于空间自身，这些钟表实际上处于静止状态，所以它们才会按照完全一样的快慢运转。正是这样的一些钟表

——其运动仅仅来自于宇宙的膨胀——才能作为同步宇宙钟来测量宇宙的年龄。

当然，要知道的是，如果你带上你的表跳到火箭上，以极快的速度横穿宇宙，那你的运动速度就会超过宇宙膨胀的速度。如果你真这样做，你的表就将按不同的快慢运转而你所得到的大爆炸时间就将是另一个结果。这样看待问题的角度当然是没问题的，但却是一个完全个人化的角度：这样一来你所测得的时间就与你的个人移动经历以及运动的状态息息相关。而当天文学家们探讨宇宙年龄的时候，他们想要的是一些普适的东西——他们寻找的是放之整个宇宙而意义不变的测量结果。整个宇宙空间变化的均匀性就提供了一种能达到这一目的的方法。[9]

事实上，微波背景辐射的均匀性为我们检验自身实际是否沿空间膨胀的方向运动提供了一种现成的方法。如你所知，尽管微波背景辐射在整个宇宙空间中具有各向同性的特点，可一旦你处于超出空间膨胀速度的运动之中，你就不会再观测到这种各向同性了。向我们疾驶而来的警车上的警笛声变得尖锐，而在飞快地离我们远去时警笛声又变得沉闷。微波背景辐射也是如此。如果你正在驾驶飞船高速飞行于宇宙空间中，那么迎面而来的微波的波峰波谷就会以一个较高的频率快速更迭，而从你的后面追过来的微波的波峰波谷的更迭频率则要低些。较高频率的微波意味着较高的辐射温度，所以你将感觉到你面前的辐射温度比你背后的辐射温度要高一些。而实际测量结果表明，在我们的地球这艘大“飞船”上，天文学家们的确发现某个方向上的微波背景辐射要热一些，而其相反方向上的微波背景辐射则要冷一些。这一事实告诉我们，不仅地球绕着太阳转，太阳绕着星系中心转，其实整个银河系都以一个超出宇宙膨胀速度的微小速度向着长蛇星座的方向运动。天文学家们只有修正了这个相对微小的速度对微波背景辐射的影响，才能清楚地看到微波背景辐射的确具有均匀性，天空中不同位置的温度非常均匀。而正是这种均匀性，不同位置之间的这种对称

性使得我们可以在描述整个宇宙的时候准确地谈论时间。

膨胀宇宙的奥妙

在我们对宇宙膨胀的阐释中有一些值得强调的细节。首先，在气球比喻中起作用的只是气球表面——两维的面（其上的每一个位置都可以用类似于地球上的经纬度来表示），而我们四下张望时看到的是三维的空间。我们之所以利用这个低维模型做例子是因为它既保留了真实情况中的本质概念，又可以形象地说明问题。我们用的是气球表面这一点你需要牢牢记住，特别是若你曾试图告诉大家气球模型中存在一个特殊点（你可能会说气球内的中心点就是个特殊点，因为气球表面所有的点都离它远去）的话。尽管你看到的这个事实是对的，但是它却对气球模型毫无意义，因为除气球表面上的点以外的任何一点都不对这一类比有任何意义。气球表面代表的就是整个空间。不在气球表面上的任何一点都只不过是这一类比的副产品，并不对应实际宇宙空间中的任何一点。[39]

第二点，如果星系所处的位置距离越远，其远离的速度就越大，那岂不就意味着距离我们足够远的星系将有可能以大于光速的速度远离我们而去？对于这个问题我们可以肯定的回答：是的。但这并不与狭义相对论矛盾。那这又是为什么呢？其中的道理又与随空间膨胀运动的钟表具有同步性有关。如我们在第3章中所强调过的，爱因斯坦证明在空间中运动的一切事物其速度都不可能超越光速。至于星系，在平均意义上，几乎不在空间中运动。星系的运动几乎可以完全归结于空间结构自身的延展。而爱因斯坦的理论并不禁止空间以一种可以驱使两点——比如两个星系——以超越光速的速度分离的方式运动。爱因斯坦的理论只限制随空间膨胀的运动被减除之后的运动速度，也就是说只限制超出空间膨胀之外的那部分运动速度。观测表明，对于沿着宇宙膨胀方向运动的星系来说，那些超出空间自身膨胀的运动速度非常有限，完全在狭义相对论所容许的范围之内，即便两个星系由于空间自身膨胀而有的分离速度超越了光速也没关系。[40]

第三点，要是空间不断膨胀，那么被拉离彼此的岂不并非只有星系？每一个星系内的空间膨胀也会使所有的恒星远离彼此，而每一个恒星内的空间膨胀，每一个行星内的空间膨胀，你我甚至世间万物内的空间膨胀，岂不会使构成各种事物的原子彼此远离？而每一个原子内的亚原子物质岂不也会被驱动着彼此远离？简而言之，空间的膨胀是不是使包括我们用的米尺在内的世间万物全部变大，从而使得我们根本无法知晓膨胀是否实际发生了呢？答案是否定的。再回想一下气球上的硬币模型。当气球表面膨胀的时候，所有的硬币远离彼此，但是这些硬币自身却没有膨胀。当然，要是我们通过在气球表面用黑笔画圈来代表星系的话，则随着气球的膨胀，那些黑圈也都变大了。但是真正抓住问题本质的是硬币而不是黑圈。每一枚硬币之所以保持大小不变是因为将锌原子和铜原子捏合到一起的力远远强于硬币胶黏其上的气球膨胀所产生的张力。与之类似，将独立的原子捏合到一起的核力，将你的骨头和皮肤捏合到一起的电磁力，以及使行星和恒星彼此接近构成星系的万有引力，都比因空间膨胀而产生的张力强得多，所以这些事物都不会变大。只有在最大的尺度上，远远大于每一个独立星系的尺度上，空间的膨胀才不会遇到任何抵抗（不同星系间的万有引力相当微弱，因为两者之间的距离太过巨大），因而只有在超星系的尺度上空间的膨胀才会驱使事物远离彼此。

宇宙学，对称性与空间的形状

要是有人大半夜的把你从睡梦中叫醒，然后让你告诉他宇宙的形状——也就是整个空间的形状——是怎样的，朦朦胧胧的你大概会没法回答。不过即使在你醉醺醺的时候，你也知道爱因斯坦证明过空间就像橡皮泥一样，所以理论上它可以是任何形状。那么你什么时候又将怎样才能回答询问者的问题呢？我们居住在一个小行星上，这颗小行星绕着一颗毫不起眼的恒星运动，我们的太阳系不过是整个银河系边缘的一个星系，相比于其他千百万个星系没有任何特别之处。那你究竟该怎样才能对整个宇宙的形状有一个认识呢？好吧，随着困意渐

渐退散，你的头脑逐渐清醒，认识到是时候再次搬出对称性来当救兵了。

如果你愿意采纳科学家们广泛持有的信念：在大尺度上，宇宙中所有的位置和所有的方向都是相对于彼此对称的，那你就很好地回答了询问者的问题。理由是，差不多所有的形状都不可能满足这一对称性的标准，因为差不多任何一种形状的某个部分或区域都在基本层面上区别于其他部分或区域。梨形上窄下宽；鸡蛋形两头尖中间粗。这些形状，虽然也具有某些对称性，但都不具有完全的对称性。将这些特别的形状排除，把视野投向那些每个区域每个方向都彼此类似的形状上，你就会发现还没被淘汰的已经出奇的少了。

我们曾经遇到过一个满足这些条件的形状。气球的球形对于在其膨胀的表面上建立所有的林肯像之间的对称性非常关键，故而这一形状的三维版本，所谓的三维球面，就是一个空间形状的候选者。但它并非是唯一一个能实现完全对称性的形状。我们继续利用易于可视化的二维模型来促进思考，想象一个无限宽无限长的橡胶薄片——完全未弯曲——其表面粘有等间距放置的硬币。随着整张薄片的扩张，我们再次得到了完整的对称性并且与哈勃的实验观测再次符合：每一位林肯都会看到其周围的林肯远离他而去，并且速度正比于距离，如图8.4所示。因而，这一形状的三维版本——想象一大块正在膨胀的透明橡胶做成的立方体，其中均匀地铺洒着星系——就是另一个可能的空间形状（如果你偏爱拿厨房里的家什作比喻，那还是想象之前提过的带瓤松饼的无限大版本，这个松饼像是个立方体，只不过要不停的无限膨胀，其中的瓤扮演星系的角色。开始烘烤后，生面团变大，使得每一个瓤离彼此越来越远）。这一形状被称为平直空间，因为其不同于之前的球形：它并没有任何弯曲（这里的“平直”是数学家和物理学家所使用的意义，并非我们平常口头上那种“平底锅”中的“平”）。[11]

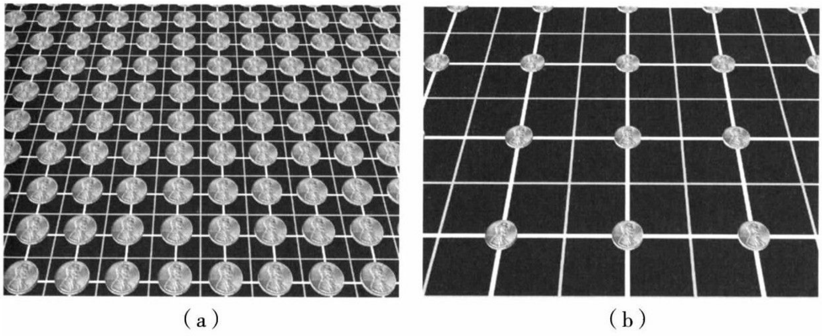



图8.4　（a）：无限平面上的每一个硬币看到的景象都与其他的硬币看到的一样。（b）：（a）中的两枚硬币相隔越远，平面扩张时其间隔增加的就越大。

球形和无限大平面形状的好处之一在于你可以沿着它们无穷无尽地走下去而不用担心到达边界。这一点非常不错因为它能使我们避开一个非常棘手的问题：空间的边界之外有些什么？如果你走进空间的边界会发生什么？如果空间没有边界，那这些问题就没有意义。但我们需要知道上述的两种形状是通过不同的方式来实现这一极具吸引力的特性的。如果你在一个球形空间一直走下去，你就会发现自己就像麦哲伦，早晚会回到起始点，永远都不会碰到边界。相反地，如果你是在无限大的平面上一直走下去，你会发现自己像电动兔[41]，永远不停，永远不会碰到边界，可是也永远无法回到起始位置。虽然这一点看起来像是弯曲和平直的形状在几何上的根本性差异，但是只要对平直空间做一些变化就会发现它将在这点上极其惊人的类似于球形。

为了形象化一点，让我们回想一下某些电子游戏，这种游戏看起来在屏幕上有边界，但实际没有边界的，因为你不能掉出屏幕：一旦你在屏幕右边的边界消失，你就会立即出现在屏幕左边的边界；如果你在屏幕上边界消失，你就会立即出现在屏幕下边的边界。屏幕是“卷在一起的”，虽然区分了上下左右，并使整个屏幕平直（未弯曲）且有有限尺寸，但却没有边界。数学上，这种形状被称为二维圆环，如图8.5（a）所示。[12]这一形状的三维版本——三维圆环——可以作为空间结构的另一种可能形状。你可以将这一形状想象成沿着3

个维度蜷曲缠绕的巨大立方体：若你在这个立方体的顶部走到尽头你就来到了底部，往后走到头就来到了前面，往左走到尽头就来到了右边，如图8.5（b）所示。这样的形状是平直的——再次提醒，这里指的平直是非弯曲的意思，不是平底锅那种平直——三维，在所有的方向上都是有限大小的，而且没有边界。

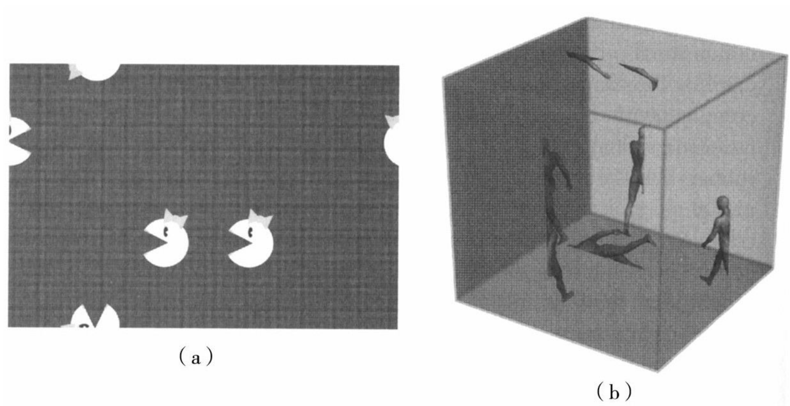



图8.5 （a）：游戏机屏幕平坦（不弯曲）且有限，但是其中的画面却没有边界，因为游戏画面“蜷在一起”。数学上，这样的形状称为二维环面。（b）：相同形状的三维版本，称为三维环面；同样的，平坦且体积有限，只不过没有边界，因为蜷曲起来。你穿过一面就会从另一面出来。

如果用膨胀空间的对称性来解释哈勃的实验观测，则空间的可能形状除了上述的这些外，还有另外一种。同三维球面那个例子一样，这个形状也很难在三维空间中画出，不过我们也可以用其两维替身——你可以把它当成品客薯片的无限大版本——来说明问题。这种形状叫做马鞍面，它是一种反球面：球面是高度对称的向外膨胀，马鞍面则是对称的向内凹陷，如图8.6所示。这里我们用点数学术语：球面具有正曲率（向外膨胀），马鞍面具有负曲率（向内凹陷），而平直空间——不管是无限大还是有限大小——则无曲率（既不向外膨胀也不向内凹陷）。[42]

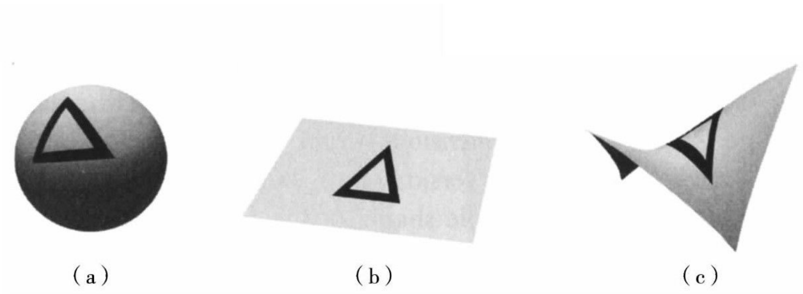



图8.6　用二维类比一下空间，一共有3种完全对称——也就是说，每一个位置看到的景象都与其他位置看到的一样——的弯曲，分别为：（a）：正曲率弯曲，一致向外鼓，举例来说就是球面；（b）：零曲率弯曲，没有任何弯曲，例如无限大平面或者游戏机屏幕；（C）：负曲率弯曲，全部向里弯，例如马鞍面。

研究人员已经证明，正曲率、负曲率和零曲率已经穷尽了能满足对称性要求——所有位置之间具有对称性；所有方向之间也具有对称性——的所有可能曲率。而这实在令人吃惊。我们讨论的可是整个宇宙的形状，这本该有无限种可能性。但是，借助于对称性的强大威力，研究人员排除了绝大部分的可能性。所以，如果你允许对称性为你引路，而那个午夜来访的提问者又同意你猜猜仅有的几个答案的话，你就有可能应付得了他的挑战。[13]

不过你可能还是想知道，关于空间结构的形状这一问题，我们为什么会得到几种不同的答案呢？我们生活在一个宇宙中，为什么不能明确它究竟是哪种形状呢？好吧，我们前面所列的形状是仅有的能与我们的信念——我们相信每一个观测者，不管他处于宇宙中的哪个位置，在最大的尺度上看到的宇宙都应当是一样的——自洽的形状。但是对称性的这种思考，尽管挑选出了少数几个选项，却不能得到最终的唯一答案。要想得到那唯一的答案，我们还需要爱因斯坦的广义相对论。

爱因斯坦方程将宇宙中的所有物质与能量（这里还要出于对称性的考虑而假定这些物质和能量均匀分布）作为输入，得到的是空间的曲率。这里的困难之处在于，天文学家们用了几十年都无法最终确定

宇宙中的物质和能量实际有多少。如果宇宙中所有的物质和能量均匀地分布于整个太空，而且其密度大于所谓的临界密度，即每一立方米中0.00000000000000000000001（$10^{-23}$）克[43]——每一立方米中5个氢原子——从爱因斯坦方程中得到的空间曲率将为正数；若宇宙中物质和能量的密度小于临界密度，则将从爱因斯坦方程中得出负曲率；若正好等于临界密度，则爱因斯坦方程告诉我们空间没有整体曲率。这一观测问题目前还没能得到确定的答案，但是目前最好的数据倾向于认为空间无曲率——也就是说实际上宇宙是平直的（但电动兔到底会不会朝着一个方向一直走下去并消失在黑暗中，又或者某天突然南辕北辙地绕到你背后——空间会不会一直膨胀下去或者会不会像电子游戏的例子那样蜷曲成首尾衔接——这样问题的答案仍然没有定论）。[14]

即便这样，就算我们不能对宇宙的形状给出一个最终的答案，我们也已很清楚地看到，我们之所以在将整个宇宙视作一个整体的时候也可以理解空间和时间，正是因为有了对称性这一核心要素的帮助。要是没有对称性的强力帮助，我们将举步维艰。

宇宙学与时空

现在我们可以将膨胀空间的概念与第3章中讨论过的时空的面包片描述联系起来以说明宇宙的历史。还记得吗？在面包片描述中，每一片面包——即使是两维的——都代表着一个特别观测者的视角下某一时间点上的三维空间。不同的观测者，根据其相对运动的不同，按照不同的角度切面包。在前面遇到的例子中，我们并没有考虑空间的膨胀；相反，我们将宇宙的结构想象为固定且不随时间改变的。现在我们要将宇宙的演化也考虑进去以便更好地探讨之前的那些例子。

为了达到这个目的，我们采用相对于空间静止的观测者——也就是说，观测者的唯一运动来自于宇宙的膨胀，就像气球上的林肯像——的视角。再次指出，即使这些观测者相对于彼此运动，他们彼此

之间还是有对称性的——他们的表显示相同的时间——因而他们以相同的方式切割时空片。在这种条件下，仅当他们的相对运动速度超过空间膨胀的速度，并且他们彼此在空间中的相对运动与空间膨胀导致的运动相反的时候，这些观测者的表才会变得不一致，导致他们的时空片的角度不再一样。我们还需指明的是空间的形状，出于对比的考虑，我们将考虑上面讨论过的可能性。

最容易画出的例子是平直的有限形状，就像电子游戏那样。在图8.7（a）中，我们给出了宇宙中的一片，你需要将该示意图看成是此时此刻的整个太空。简单起见，我们的银河系被画在图的中心，但你需要记住这并不表示我们的银河系有何特别之处，宇宙中没有任何位置有特殊的地位。图中的边界并不真正存在。图的上端并不就是宇宙的边界，你迈过最上端时将会回到最下端。与之类似，图的最左端也不是宇宙的尽头，迈过最左端你将回到最右端。而要想令这幅图符合天文学观测，我们还需要将图的每条边都从其中心点开始各向两边延伸至少140亿光年（差不多850亿兆千米），甚至更长也是可能的。

反之，我们抬头仰望漆黑清澈的夜空时看到的种种光亮都是很久以前——数百万年甚至上亿年以前——即已发射出来的光。这些光经过漫长的旅程，直到今日才到达我们这里，进入我们的天文望远镜中，使我们可以通过它们感受外太空的神奇景观。因为宇宙一直在膨胀，所以在这些光束刚刚射出的远古时代，宇宙比之今日要小得多。我们通过图8.7（b）来展示这一点。在这张图中，我们将现在的时间片放在最右端，从右至左的时间片代表的就是我们的宇宙在越来越早的时期的样子。如你所见，宇宙所处的时期越早，其整体尺度以及星系之间的间距就越小。

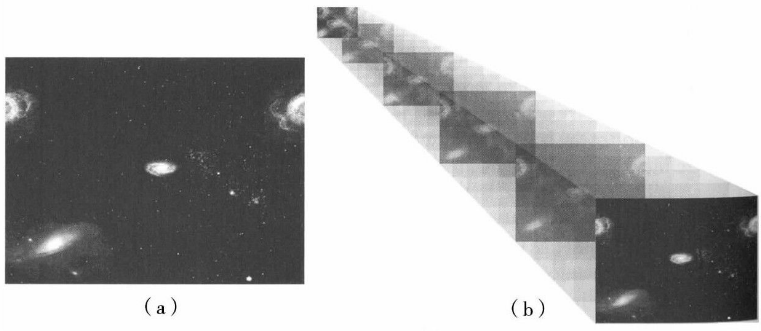


图8.7　（a）：现在的所有空间的示意图，假定了空间平坦且有限，也就是说看起来像游戏机屏幕。注意上右的星系绕回到上左。（b）：所有空间随时间演化的示意图，我们把时间分片以便看起来清楚些。注意到空间的整体尺寸核星系的间隔随着时间回溯而减小。

在图8.8中，你看到的是一束光的历史，这束光从遥远的，或许100亿光年外的河外星系射出，向着银河系中的我们飞来。在图8.8（a）的第一片中，这束光刚刚射出；从其后的那些片中我们可以看到，即使宇宙越变越大，这束光也照样离我们越来越近，最后来到我们跟前，如最右边的时间片所示的那样。在图8.8（b）中，我们将每一时间片中该光束的前端连接起来，就得到了这束光穿越时空的路径。因为我们可以从很多不同的方向接收到光波信号，所以我们另用一张图片[图8.8（c）]展示来自于不同光源的光束在时空中留下的轨迹。

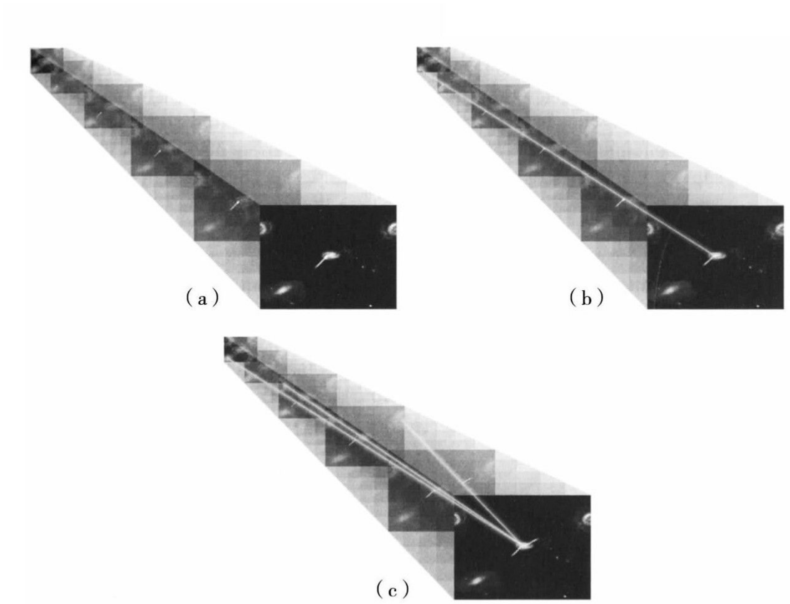



图8.8　（a）：很久以前从远处发出的光正接近银河系。（b）：我们最终看到远方的星系时，我们看到的是穿越了空间和时间的远方星系，因为远方星系的光是在很久以前发出来的。高亮显示的是这束光在时空中的路径。（C）：我们今天所见的来自于各种天体的光在时空中的路径。

这些图片戏剧性地说明了为什么来自太空的光束能被用作封存宇宙时间的胶囊。当我们望向仙女座星系时，我们看到的光发自于300万年以前，所以我们看到的实际上是仙女座星系过去的样子。当我们望向后发座星系团时，我们看到的光发自于3亿年以前，所以我们看到的后发座星系团比看到的仙女座星系还要老。即使这个星系团中所有星系中的所有恒星此刻都一下子变成了超新星，我们所能看到的也是没有任何突变的景象，并且在接下来的3亿年间我们也不会看到它们的集体爆发；只有在3亿年后，超新星爆发时发出的光到达我们这里我们才能了解当时发生的一切。与之类似，要是现在时间片上的后发座星系团中的一位天文学家正在用一台超级天文望远镜探看我们地球，她所看到的也只会是大量的蕨类植物、节肢动物以及远古爬虫；她绝不会看到中国的万里长城或是巴黎的埃菲尔铁塔，要想看到这些，她还得再等3亿年。当然，这位天文学家想必已受过专门的宇宙

学培养，明白她所看到的光源来自于3亿年前的地球，并且将它观测到的地球早期细菌知识安置于她自己的宇宙时空条的适当时期，适当的时间片上。

上述的一切都预先假定我们以及后发座星系团中的天文学家仅随同空间的膨胀而运动，因为这一假定保证了她从时空条中切得的那片与我们的一致——即保证了我们与她对现在的认识具有一致性。不过，要是她不再跟我们同步，而是以大于宇宙膨胀的速度穿行于太空，那么她的时间片就会相对于我们倾斜，如图8.9所示。在这种情况下，就像在第5章时我们同楚巴卡一道发现的那样，这位天文学家的现在就会同我们所认为的过去或者未来（究竟是过去还是未来要看到底是向着我们运动还是远离我们运动）保持一致。需要注意的是，这样一来，她的时间片就不再具有空间上的各向同性。图8.9中每一个倾斜的时间片所描述的宇宙是一个包括一段不同时间点的宇宙，因而时间片不再具有均匀性。这样一来，我们在描述宇宙历史时的复杂程度就会增加很多，而正是因为这样，物理学家和天文学家一般不愿采用这样的分析视角。一般来说，物理学家和天文学家采用的是仅随着宇宙膨胀而运动的观测者视角，这样一来所有的时间片都能保有各向同性——但从根本上讲，所有的视角都一样有效。

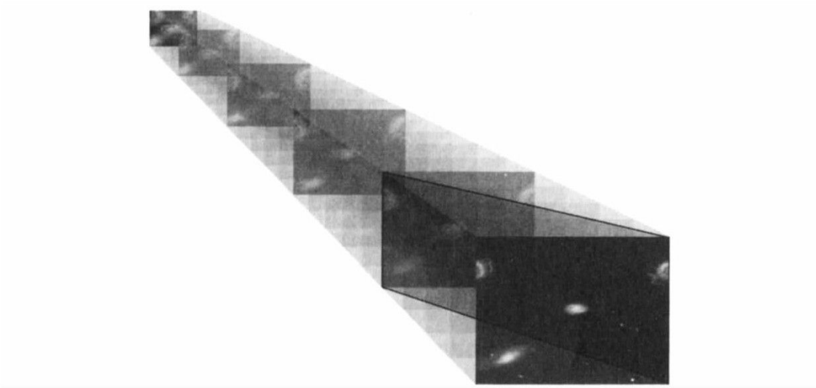


图8.9　一个超越宇宙空间扩张速度的观测者的时间片段。

沿着宇宙时空条的左边望去，我们会发现宇宙变得越来越小，越来越密。当我们往自行车车胎中不断地打气时，车胎就会变得越来越热，而宇宙也是如此，当空间不断缩小，物质和辐射变得越来越密的时候，整个宇宙也就变得越来越热。如果我们追溯到宇宙诞生后的百万分之一秒，我们将发现宇宙如此之密而且如此之热以至于普通物质都分解成由大自然中的基本粒子所构成的原初等离子体。而当我们继续追溯，直到接近时间为零的时刻——大爆炸的那一刻——整个的已知宇宙都将被压缩到一个小到难以想象的尺寸上，以至于相比于当时的宇宙，这句话结尾处的句号都是真正的庞然大物。由于早期的密度实在太过惊人，而且当时的物理条件又是那么极端，因此现代所能拥有的最好的物理理论都无法告诉我们当时的情况。出于一些我们将要在后面详加介绍的理由，20世纪发展出来的那些已经取得了巨大成就的物理定律在如此恶劣的条件下不再有效，使我们在时间的源头处失去了方向舵。我们即将看到，近年来的一些进展为我们点燃了希望的灯塔，不过即便如此，现在我们也只能承认自己对宇宙时空条最左端的那片起始区域认识并不完整，只能模模糊糊地将它画在那里——它就是我们的旧地图上的未知区域。最后我们用图8.10来给出一幅粗线条的宇宙历史。

其他形状

到目前为止，我们一直假定宇宙空间具有像电子游戏屏幕那样的形状。对于其他的形状，我们也将得到一些相同的性质。比方说，如果实验数据最终证明空间为球面形状，那么当我们沿着时间追本溯源的时候就会发现，球面的尺寸变得越小，宇宙就变得越热越密。最后，在时间的尽头我们就遇到了某种大爆炸的起点。要想画出一幅图8.10那样的示意图可是一项很有挑战性的工作，因为球面很难被简单的一个挨一个的摆好（不信你可以试着想象一个“球形面包切片”，它的每一切片都是一个球面，罩在上一个切面的外面）。不过除了画图上的这点困难，重要的物理部分都与我们前面的讨论一样。

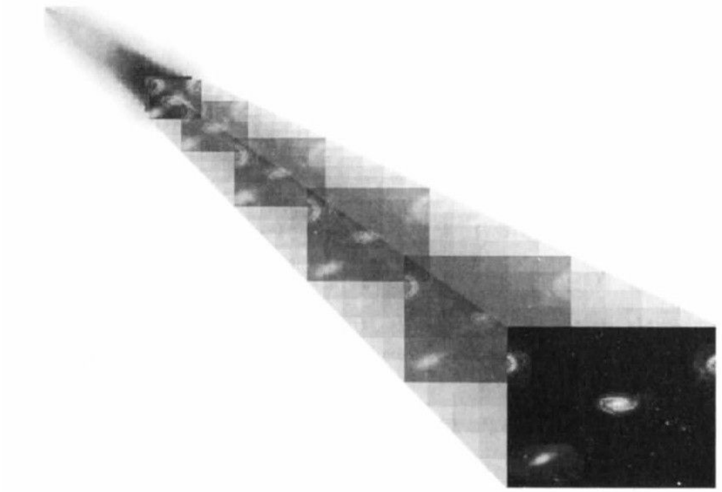


图8.10　平坦且有限的宇宙的历史——时空“片”。上面之所以模模糊糊是因为我们对宇宙的开端一无所知。

无限大平直空间与无限大马鞍面也具有一些共同的性质，不过两者有一本质区别。来看一下图8.11，其中代表平直空间的时间片可以无限延展（当然，我们画出的只是其中的一部分）。我们所研究的时间越早，空间就越小。所以在图8.11（b）上我们可以看到，时间越是推向过去，星系就变得越密。但是，空间的整体大小却保持不变，这又是为什么呢？这个嘛，无限大是一件非常古怪的事。如果空间是无限大，那么你把它缩小2倍的话，空间就是原来的1/2，也就是1/2无限大，可是1/2无限大也是无限大。所以，当你逆着时间一路向过去探查，你会发现世间万物全都彼此靠近，密度也变得越来越大，但是宇宙的整个大小却仍是无限大；在无限大的空间内一切都变得越来越近。这将使我们得到一幅全然不同的大爆炸图像。

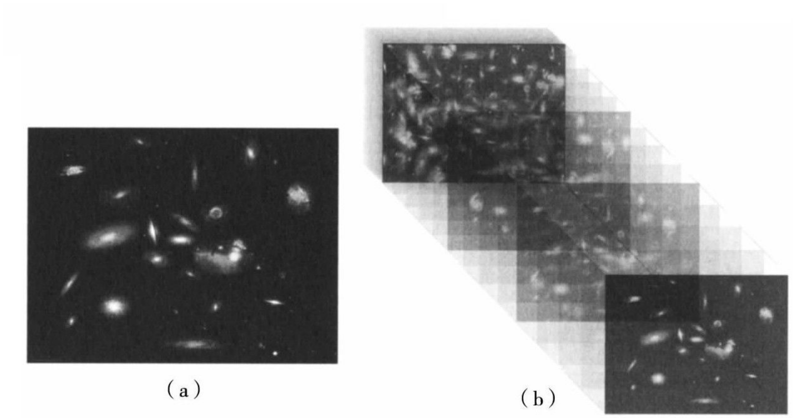


图8.11　（a）：无限空间的图释，其中有大量星系。（b）：更早的时期星系收缩——所以星系离得更近且更加密集——无限空间的整体大小还是无限。我们对较早时期发生了什么一无所知，因而只好还用些模糊的片代表，只不过这里都是无限空间。

一般来说，我们将宇宙想象为诞生于一个点，大致如图8.10所示，这个点之外并没有空间和时间。然后，从某次突然的爆炸开始，空间和时间就开始摆脱它们的压缩形态，宇宙就开始膨胀。但如果宇宙的空间是无限大，那么在大爆炸的那一刻就已经存在了一个无限大的空间区域。在这原初的一刻，能量密度高涨，温度高得不可想象，但是这种极端条件无所不在，并不是只存在于某一点。按这种说法，大爆炸并不是发生于某一点，而是爆发于整个无限大的空间范围内的所有地方。将此与传统意义上的发源于一点的说法相比，这种大爆炸就相当于说有很多的大爆炸，无限大的空间范围内的每一点上都有大爆炸。大爆炸之后，空间膨胀，但是其整体大小却不可能发生变化，因为一个无限大的东西是不可能变得更大的。那么，为什么当你从图8.11（b）的左边往右边望去的时候，会发现星系（一旦它们形成）之类的事物之间的距离变大了呢？所有的观测者，无论你我还是其他的什么人，都会发现围绕着自己的星系正在远去，就像哈勃所发现的那样。

必须牢记的是，无限大平直空间并不仅仅是一个只具有学术上意义的例子。我们将在后面的章节中看到，存在着空间的整体形状并不

是弯曲形状的坚实证据；而且，也没有证据表明空间为电子游戏屏幕那样的形状，因而我们可以说平直，无限大的空间形状真的是时空大尺度结构的有力竞争者。

宇宙学与对称性

对于对称性的思考显然已成为现代宇宙学发展中不可或缺的部分。时间的意义，把整个宇宙当成一个整体来研究时的时间概念是否可用，空间的整体形状，甚至是广义相对论的深层次理论框架，所有的这一切都在根本上依赖于对称性。即便如此，我们还将发现存在另外一种从对称性中获知宇宙演化的方法。通过对宇宙温度史的学习，我们将会发现，从大爆炸刚刚过去之后的极热时刻一直到我们测量外太空所得到的只有绝对温度几度的今天，整个的这段过程都可以从对宇宙温度的研究中了解到。而且，正如我即将在下一章中要阐释的，由于热与对称性之间的相互依赖关系，我们今日所看到的对称性很可能是远远丰富于今日的对称性的残存部分，而正是这些更为丰富的对称性铸就了早期宇宙并且决定了有关宇宙的某些我们最熟悉且具有本质意义的性质。

## 第9章　蒸发真空

热，一无所有，统一

在整个宇宙历史中，95%之多的时间内，我们都可以这样简略地描述我们的宇宙：宇宙在膨胀。由于膨胀，物质继续扩散。宇宙的密度持续减小，温度继续降低。在最大的尺度上，宇宙具有对称、各向同性的外观。不过宇宙并不总是如此。关于最早时期，我们会有一段气氛热烈的报道，因为在那段时间里，宇宙在极速改变。我们现在知道，宇宙过去所发生的一切直接影响着我们今天对它的认识。

在本章中，我们将焦点集中在大爆炸之后的最初时刻。人们相

信，宇宙的对称性在那个时候发生过巨变，而每一种改变都在宇宙的历史中留下了一段意义截然不同于其他的时期。尽管关于今日宇宙的报道可以参照几十亿年前的内容，但是报道宇宙中对称性极速变化的初创时期的工作则非常富有挑战性，因为我们完全不熟悉那个时候物质与力的基本结构。我们要了解的内容与热和对称性的相互影响有关，并且我们需要反思一下我们关于真空和一无所有的概念。我们将会看到，这样的反思将并不仅仅丰富我们关于宇宙最初时刻的知识，还能够引导我们向着牛顿、麦克斯韦，特别是爱因斯坦追寻的统一之梦迈进一步。同样重要的是，这样的一些理论进展将为最现代的宇宙学理论框架暴涨宇宙学——一种能够回答一些标准大爆炸宇宙学解释不了的最困难最麻烦问题的方法——打下基础。

热与对称性

事物在极热或极冷的时候一般会有所改变。有的时候这种改变太过强烈，以至于你无法认出它的本来面貌。大爆炸之后处于极热状态的宇宙随着空间的膨胀和冷却而温度降低，了解温度改变所带来的效应将会帮助我们搞清宇宙的早期历史。但是真正开始之前，让我们先看看比较简单的例子，我们来了解一下冰。

如果你将一块非常冷的冰加热，你会发现最初没什么变化。虽然那个时候冰的温度在增加，但它看起来一点变化都没有。但是如果一直把它加热到0摄氏度，并且不停下来，你会突然看到发生了某些奇妙的事情。固态的冰开始融化成液态的水。不过这种熟悉的现象不是你要注意的，你要注意的是冰和水之间的密切关系。一种是坚硬的冰，而另一种是流动的水。只从表面看的话，我们很难想象它们的分子结构竟然都是$H_2O$。如果你以前从没见过冰和水出现在一起，你可能就会认为它们是完全无关的。那么在温度达到0摄氏度的时候，你就会非常吃惊地发现，它们可以彼此转化。

继续加热水，你又会发现除了温度升高外没有任何其他变化。等

到温度达到100摄氏度的时候，又会突然出现一种变化：液态的水开始沸腾并且变成了水蒸气。表面看来，这种热得发烫的气体同液态的水和固态的冰没有任何关系。但是我们知道，这3种物质具有相同的分子组成。从固态到液态以及从液态到固态的这种变化称为相变。如果温度变化的范围足够大[1]，大部分物质都会有类似的改变过程。

对称性在相变的过程中具有核心地位。在几乎所有的情况中，我们都会发现相变前后某种东西的对称性发生了变化。比如，在分子的尺度上，冰具有晶体结构，所有的$H_2O$分子排成有序的六边形。就像图8.1中的盒子一样，冰的分子也会在某些特定的操作下保持整体不变，比如绕着某些特定的轴转动60度这样的操作。另一方面，如果我们加热冰的话，其晶体结构就会化为混乱但却均匀的分子团——液态的水——这样的结构在关于任何轴的任何角度的变换下都保持不变。因而，将冰加热使其经历固液相变，我们就可以使它具有更高的对称性（只靠直觉的话，你可能会认为更加有序的东西，比如冰，会具有更高的对称性，但事实并不是这样；如果我们说某种东西具有更高的对称性，那么这种东西一定能在更多的变换下——比如转动变换下——具有不变性）。

类似地，如果我们加热水使其变为水蒸气，期间的相变也会导致对称性增加。水中的$H_2O$分子紧紧挤在一起，每个分子氢的一面紧紧挨着其邻近氧的一面。如果你能够旋转这个或那个分子，你就会破坏水中的分子排列模式。但是当水沸腾变为水蒸气的时候，分子就会四处飘散；$H_2O$分子的排列方向就再也不具有任何规律可言，这个时候不管你怎样旋转单个分子或是一群分子，水蒸气表面上都没有任何改变。也就是说，固液相变能够增加对称性，液气相变也会增加对称性。对于大部分的物质（不过并不是所有物质[2]）来说，也是如此。当它们经历固液相变或液气相变的时候，对称性会相应的增加。

反过来，冷却水的时候也会有类似的过程发生，只不过方向要相反。例如，冷却水蒸气的时候，最开始什么变化也没有，但随着温度

低于100摄氏度，水蒸气突然开始凝结成水；继续冷却水，仍然没有任何现象，直到温度低于0摄氏度，水突然开始结冰。还像前面那样考虑对称性的话——只不过一切相反——我们就会得出结论：在这样的两种相变过程中，对称性减少了。[44]

关于冰、水、水蒸气及其对称性，我们能说的就是这么多。那么它们与宇宙又有什么关系呢？我们现在就来谈谈。20世纪70年代的时候，物理学家们认识到并不只有宇宙中的物质才能经历相变，宇宙本身作为一个整体也可以发生相变。在过去的140亿年间，宇宙稳步地膨胀减压，就像一个压力减少的车胎会慢慢变凉；膨胀中的宇宙，它的温度也在稳定地降低。在这个温度降低的过程中，大部分的时间内什么特别的事情也没有发生。但是我们有理由相信，宇宙曾经历过某些特别的临界温度——就像水蒸气的100摄氏度和水的0摄氏度——在这样的临界点处宇宙急剧改变，对称性快速减少。很多物理学家相信我们现在生活在宇宙“凝聚”或“冻结”的相，现今的宇宙与早前的宇宙应该有很大的不同。说到宇宙的相变时，我们不能简单地把它理解成气体凝结成液体、液体冻结成固体的过程；虽然定性来说，宇宙的相变与这些人们熟知的相变过程有些类似。我们要了解的是，宇宙在某个温度处经历相变的时候，发生凝结或冻结的“物质”不是别的，而是场——更确切地说，是希格斯场。我们马上就看看这究竟是怎么一回事。

力，物质，希格斯场

场的概念是很多现代物理学的理论框架。第3章中讨论过的电磁场或许是所有自然界中的场最简单最为人所熟知的一个。想想无线电、电视信号、手机沟通、太阳的热与光，我们所有的人都在电磁场的海浪中畅游。光子是电磁场的基本组成，我们可以将它看做是电磁力的微观传递者。当你看到某个东西的时候，你可以这样想象：波动的电磁场正在进入你的眼睛并且刺激着你的视网膜；又或者说是大量的光子正进入你的眼睛并且刺激着你的视网膜。正因为如此，光子有

的时候被称为电磁力的信使粒子。

引力场是另一种大家熟悉的场，引力场无时不在，始终如一地把我们和我们身边的事物牢牢地锚定在地球表面。和电磁场一样，引力场的波涛也把我们深深地浸入其中。在我们看来，地球的引力场起主要作用，但同时我们也能感受到太阳、月亮和其他星球的引力场带来的影响。正如光子组成了电磁场，物理学家们也相信引力场是由引力子组成的。虽然实验上还没能发现引力子，但是人们对此并不惊奇。引力是人们目前为止所发现的最弱的力（一块普通的电冰箱磁铁就能吸起来一个纸夹，也就是说这块普通的磁铁施加给纸夹的电磁力就能大过纸夹所受到的引力，可见引力是多么的弱），因而我们能够理解实验学家们还没能发现这种最弱的力的组成粒子。但即使这样，很多物理学家仍然相信引力子（引力的信使粒子）能像光子（电磁力的信使粒子）传递电磁力一样传递引力。你失手落下一只杯子的时候，你可以想象成是地球的引力场在推着这只杯子向下飞落；或者，你也可以用爱因斯坦更加精练的几何描述：地球的存在导致了时空结构的弯曲，杯子沿着弯曲的时空结构滑行；又或者，如果你相信引力子真的存在的话，你可以把这个过程想象成繁忙的引力子在杯子和地球之间来来回回的运动，传递着引力要“告诉”杯子向地球上落下的“信息”。

除了这两种著名的力场，自然界中还有另外两种力：强核力和弱核力，它们的影响也通过场表现出来。核力之所以不像电磁力和引力那么著名是因为它们只在原子和亚原子尺度上起作用。但即使这样，它们也影响着我们的日常生活：使太阳能够发光放热的核聚变，原子反应堆中的核裂变，以及放射性元素如铀和钚的放射性衰变等，其重要性不言而喻。强核力和弱核力的场称为杨米尔斯场，这是根据杨振宁和罗伯特·米尔斯命名的，这两位物理学家在20世纪50年代提出了研究强核力和弱核力所需要的理论基础。正如电磁场是由光子组成，引力场被认为是由引力子组成，强核力场和弱核力场也有其微粒组分。强核力的粒子被称为胶子，而弱核力的粒子是W粒子和Z粒子。

这些传递力的粒子的存在已经在20世纪70年代末80年代初由德国和瑞士的加速器实验所证实。

场的概念也可以应用于物质。量子力学的概率波自身就可被看做是遍布于空间的场，而正是这种场使得某个物质粒子按一定概率出现在某个位置。比如说，一个电子，既可以被看成是一个粒子——能够在图4.4所示的荧光屏上留下一个点——也可以（并且必须）被看成是一种波场，正是这种波场使得图4.3（b）中的荧光屏上出现相干图样。[3]事实上，尽管我并不打算进一步深入探讨，[4]但有必要提一下，电子的概率波与所谓的电子场密切相关，电子场是一种在很多方面与电磁场非常类似的场，其中电子扮演着与电磁场中的光子类似的角色，电子是电子场的最小组分。这里对场的有关描述也适用于所有其他种类的物质粒子。

讨论过物质场和力场之后，你可能会认为我们已经讨论过所有的场了。但很多人相信故事远未结束。很多物理学家坚信，还有第3种场。虽然人们从未在实验上发现过这种场，但在过去的几十年间，这种场却一直在现代宇宙思想和基本粒子物理中扮演着重要角色。这种场被称为希格斯场，以苏格兰物理学家彼得·希格斯的名字命名。[5]如果我们要在下一小节中讲到的思想是正确的，那么宇宙中就会填满希格斯海——大爆炸后的冷却残余——正是它的存在，组成你、我以及我们所遇到的一切事物的粒子才具有它们该有的很多性质。

冷却宇宙中的场

温度对场的影响同温度对物质的影响一样。温度越高，场的值就会变化得越猛烈——就像水壶中急剧沸腾的水的表面一样。无论是在今日冰冷的深层空间中（绝对零度之上2.7度，或者用惯用的表示，2.7开），还是在暖和些的地球上，场值的变化都不大。但是，在大爆炸之后的一瞬间，温度极高——在大爆炸之后的$10^{-43}$秒，人们相信那时的宇宙温度可高达$10^{32}$开——所有的场都处在剧烈动荡中。

随着宇宙渐渐膨胀，温度逐渐降低，物质及辐射最初所具有的巨大密度稳步衰减，宇宙中的大片区域变得前所未有的空荡，场的波动渐渐平息。对于大多数的场来说，这意味着它们的值总体上趋近于零。在某些时刻，某种场的值会略大于零（形成波峰）；而在另外的某些时刻，这种场的值又可能略微低于零（形成波谷），但是平均下来，大部分场的值都会逼近零——直觉上，我们认为场值在虚无的真空中就应该是零。

希格斯场就在这里登场了。研究人员已经认识到，在大爆炸之后的灼热温度下，希格斯这种特殊的场具有同其他的场类似的性质：狂乱的上下波动。但是研究者们相信（正如温度降低得足够多时，水蒸气会凝结成液态的水），随着宇宙温度降低得足够多，希格斯场会在整个空间中凝聚成一个特别的非零场值。我们在物理学中将这个称为非零希格斯场真空期望值的形成——不过我们不用技术术语，我在下面会将其称为希格斯海的形成。

这多少有点像把青蛙扔到热的铁碗中时发生的事。我们一起来看看图9.1（a），图中是一个烧热的铁碗，铁碗的中间放着一些蚯蚓。刚开始的时候，青蛙会四处乱蹦——忽上忽下，忽左忽右——以免自己的腿被烫伤。总的来说，这个时候青蛙离铁碗的中心很远，甚至都不知道那里还有美味。但是随着碗的温度降低，青蛙渐渐地冷静了下来，懒得到处乱蹦了，于是它就会慢慢滑到碗的中央。在那里，它就会发现那些蚯蚓，然后享用自己的美味了，如图9.1（b）所示。

但如果碗的形状有所不同，如图9.1（c）所示，则结果就会有所不同。我们再想象一次：首先，碗还是很热，蚯蚓还是待在碗的中心，只不过现在碗的中心是鼓起的。然后，你再把青蛙扔进碗中，它又会四处乱蹦，还是没能注意到位于碗中心的礼物。接下来，温度逐渐降低了，青蛙又懒了下来，不再跳了，慢慢地顺着光滑的碗壁滑了下来。但是因为现在的形状不一样了，青蛙到不了碗的中心了，这时不用费劲就能到达的地方是碗的谷底，而这里没有蚯蚓，如图

9.1（d）所示。

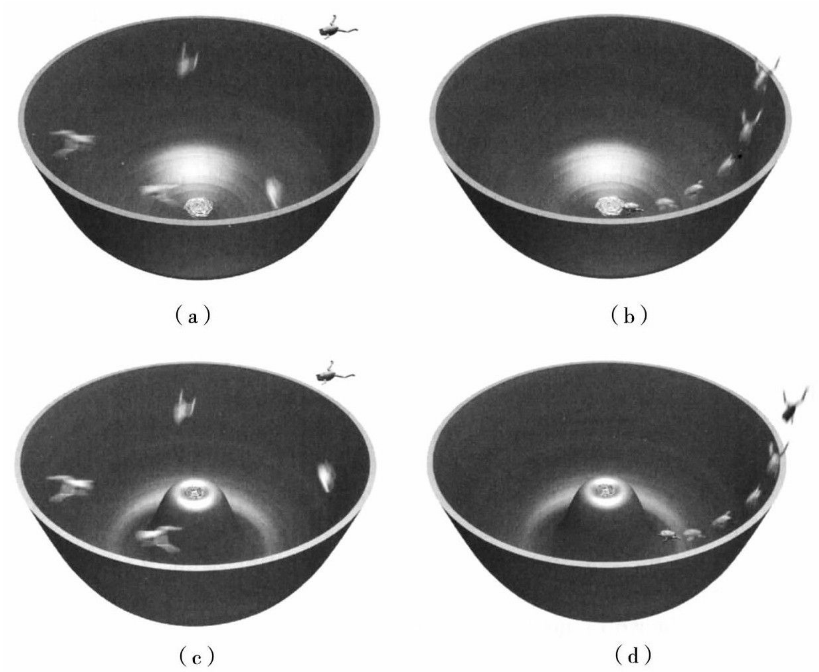



图9.1　（a）：跳到滚烫铁碗中的青蛙被烫得四处乱蹦。（b）：碗慢慢凉了下来，青蛙也就渐渐的不跳了，于是滑到了碗的中间。（C）：同（a）时的情况一样，只不过碗的样子不一样。（d）：同（b）的情况一样，只不过这次碗凉下去的时候，青蛙滑到了谷底，而这个位置离碗的中心（螺纹的地方）还有一段距离。

如果我们将青蛙与蚯蚓之间的距离想象成场的值（青蛙离蚯蚓越远，场的值也就越大），将青蛙所处的高度想象成场中所具有的能量（青蛙在碗中的位置越高，场中的能量也就越大），那么，图9.1中的这些例子就能很好地帮助我们理解当宇宙温度降低时场的行为。宇宙的温度很高的时候，场的值也在急剧变化，就像碗中四处乱蹦的青蛙一样。随着宇宙逐渐冷却下来，场也变得“冷静”多了，场的值开始减少变化，最终滑到最低能量状态的位置。

于是问题来了。就像青蛙例子一样，这里也可能有两种定性上不同的结果。如果场的能量碗——所谓的势能——的形状类似于图

9.1（a）所示的那样，整个空间中的场值都会滑落为零，也就是势能碗的中心处，就像青蛙最终会滑落到蚯蚓堆一样。但是，如果场的势能看起来像图9.1（c）所示的那样，场值就不会总能落到零，也就是场的中心。相反的，正如青蛙会滑到碗中的谷底，而那里到蚯蚓堆还有一段距离，场值也将滑落到凹处——此处与碗中心的距离非零——而这就意味着场有非零值。[6]后者即为希格斯场的特性。随着宇宙冷却下来，希格斯场的值落入谷底，永远没法变成零。又因为我们所描述的乃是整个空间中一致发生的情况，所以整个宇宙中就会均匀地填满非零的希格斯场——希格斯海。

这一过程的发生原因使希格斯海的基本特性显露无遗。当空间中的某处变得越来越冷、越来越空时——因为其中的物质与辐射越来越少——该区域的能量就会变得越来越低。将这种情况推至极限，你将会认识到，如果将你某一空间区域的能量尽可能地降低，那么这一区域就会变得尽可能的空。对于填塞空间的普通场来说，当它们的场值滑落至图9.1（b）所示的碗中央时，它们的能量就达到最低；当它们的场值为零时它们的能量也为零。这在直觉上很好理解，因为我们要想将空间中的某处变空，就得令一切事物，包括场值，都变成零。

但是对于希格斯场来说就不一样了。正如青蛙要想到图9.1（c）中的中央高地上并跟蚯蚓堆零距离的话，它就得有足够的力量从碗中的谷底蹦到高地上；希格斯场要想达到碗中央并且使其值为零的话，也得有足够的能量使其超越碗中的高地。相反的，如果青蛙没什么劲了，它就会滑到碗中的谷底，如图9.1（d）所示——这样一来距离蚯蚓堆就是非零距离了。类似地，要是一个希格斯场没有能量或者能量太少的话，该希格斯场也只能滑落到碗谷——与碗中央的距离非零——从而具有非零的场值。

为迫使希格斯场有一个非零值——看起来只有这个值才能使你将场从空间中完全除去，看起来也只有这个值才能使你获得什么也没有的空间——你必须提高它的能量，这样一来，从能量的角度说，空间

就不会达到它所能达到的空。这听起来有点自相矛盾，消除希格斯场——也就是说使其场值为零——等价于为所讨论的区域注入能量。我们可以用那种消除噪声的耳机打个不太恰当的比方。这种耳机会制造声波来抵消来自外界的噪声，从而使环境噪声没有机会敲击你的耳膜。如果耳机工作良好，那它制造声音的时候你就什么也听不见，而当你把它关掉的时候你就会听到环境噪声。正如按程序制造的声音弥漫于耳机的时候，你听到的反而少了。研究人员相信，当其间弥漫着希格斯海的时候，冰冷的空间所拥有的能量才能达到最少——空间也才能达到尽可能的空。学者将最空的空间称为真空，现在我们知道，真空中实际上满是均匀的希格斯场。

为整个空间中的希格斯场假定一个非零值——形成希格斯海——的这种程序被称为对称性自发破缺[45]，是20世纪后几十年中理论物理学界出现的最重要的思想。我们一起来看看这是为什么。

## 希格斯海与质量起源

如果希格斯场的值不为零——如果我们都在希格斯场的汪洋中畅游——我们是不是应该能够感觉到它或者看到它又或者能够用别的什么方式知道它的存在呢？当然是这样。现代理论宣称我们能够感知到希格斯场的存在。前后挥挥你的手臂，你能感觉到你的肌肉牵引着你的手臂来来回回地摆动。如果你再多握一只碗，你的肌肉就得更卖力一点才能使你挥动你的手臂，这是因为它需要移动更大的质量，因而也就需要更多的力。在这层意义上，一个物体的质量代表的是它抗拒被移动的能力；或更确切地说，一个物体的质量代表的是它抗拒运动状态的改变的能力。所谓运动状态的改变就是获得加速度，也就是被改变速度（不管是大小还是方向）；比如开始向左运动，然后又向右运动，接着再向左运动。但是，这种抗拒被加速的能力来自于哪里？或者，用物理学的语言说，是什么使得物体具有惯性？

在第2章和第3章中，我们看到了牛顿、马赫和爱因斯坦各自提出

的种种回答，这些回答都只是这一问题的部分解答。这些物理学家们都试图找到一种静止的标准，从而使得人们可以相对于这种静止来定义加速度，他们之所以讨论旋转水桶这个实验为的就是这个目的。对于牛顿来说，这一标准就是绝对空间；对于马赫来说，这一标准就是远方的群星；对于爱因斯坦来说，开始的时候这一标准是绝对时空（狭义相对论），之后换成了引力场（广义相对论）。但是，一旦确定下来这一静止的标准，特别是确定下了定义加速度的基准，这些物理学家就再也不深入探讨物体为什么会抗拒加速度了。也就是说，这些科学家中没有一个讨论过物体获得质量——也就是惯性，抗拒被加速的性质——的机制。现在我们有了希格斯场这一概念，物理学家们可以对这一问题给出答案了。

组成你手臂的原子，你捡起来的保龄球，都是由质子、中子和电子构成的。而实验学家们又在20世纪60年代末期发现，质子和中子是由3个更小的粒子组成，这种更小的粒子就是夸克。所以，当你挥动手臂的时候，你实际上在挥动组成手臂的夸克和电子，这样的观念将使我们接近事情的关键之处。现代理论所说的、我们都深浸其中的希格斯场，与夸克和电子都可以相互作用：这种相互作用使希格斯场阻碍了夸克和电子的加速，就好像掉进蜜罐的乒乓球会被粘得难以移动。正是这种阻力拖住了各种微粒的运动，从而使你能够感受到你自己手臂的质量以及你所挥舞的保龄球的质量，又或者你正在投出的其他物体的质量，甚至是你加速冲过百米终点时你整个身体的质量。因而，我们的确能感知到希格斯海的存在。我们每天使用数千次、用以改变这样或那样物体运动状态——通过施加加速度——的力都是为了抗拒希格斯海的阻力而存在的。[8]

蜜罐的这个比喻贴切地抓住了希格斯海的某些方面。你要是想给一个浸在蜜罐中的乒乓球加速的话，你就得付出比在乒乓球台上更大的力气才行——蜜罐中的乒乓球比不在蜜罐中的乒乓球更加抗拒速度的改变，看起来就好像乒乓球的质量变大了一样。希格斯海的情形与

此类似：由于基本粒子与无所不在的希格斯海之间存在相互作用，基本粒子也会抗拒自身速度的改变——换种说法就是获得了质量。不过，我必须提醒读者注意的是，蜜罐的这个比喻有3处带有误导性质的地方。

首先，你总可以伸手到蜜罐里把乒乓球拿出来。这样，它那种抗拒加速的特点就消失了。但是基本粒子就不一样了。今天，人们普遍认为希格斯海无处不在；因而，我们没办法使粒子逃脱它的影响；无论什么地方的粒子都有它们本该有的质量。其次，蜜罐使得乒乓球抗拒所有的运动；但是受希格斯海影响的基本粒子只抗拒加速运动。一个在蜜罐中运动的乒乓球速度会慢慢减小；但是一个在外层空间匀速运动的粒子则不会因为希格斯海的“阻力”而速度减小，它仍旧保持匀速运动。只有当我们加速或减速一个粒子的时候，希格斯海才会通过我们所施加的力来展现它的存在。最后，考虑由基本粒子所构成的各种现实物质的时候，情况又有所不同，质量还有另外的来源。强核力将夸克结合在一起，组成质子和中子。也就是说，在夸克之间来来往往的胶子（强核力的信使粒子）将夸克“胶连”在一起。人们通过实验发现，胶子通常具有很高的能量。根据爱因斯坦的质能方程E＝mc2，我们知道，能量（E）可以以质量（m）的形式体现出来。因而，质子和中子内部的胶子可以为这些粒子的总质量带来很大的贡献。总而言之，更准确的说法是：希格斯海那蜜罐一般的黏滞力使电子和夸克这样的基本粒子获得质量；这些基本粒子又可以组合成复合粒子，比如说质子、中子以及原子之类；但是在这个组合过程中，其他的质量起源也要起作用。

物理学家们假定了希格斯海阻止不同种类粒子加速的能力有所不同。这一假定很重要，因为已知的各种基本粒子质量各不相同。比如，人们早就知道质子和中子是由两种夸克（分别称为上夸克和下夸克：质子由两个上夸克和一个下夸克组成；中子由两个下夸克和一个上夸克组成）组成的，多年的原子对撞机实验使人们又发现了另外的

4种夸克。所有这些夸克的质量分布在一个很大的范围内——最小的只有质子质量的0.0047倍，而最大的则有质子质量的189倍。物理学家们认为，各种基本粒子的质量之所以彼此不同，是因为不同种类的基本粒子同希格斯海的相互作用强度各不相同。如果一个粒子同希格斯海的相互作用很小甚至根本没有，那它就只受很小或根本不受希格斯海的阻碍，因而其质量也就很小或者根本没有。光子就是一个这样的粒子，光子可以完全不受影响的穿行在希格斯海中，因而也就全无质量。另一方面，如果一个粒子与希格斯海的相互作用非常的强，那它就会有很大的质量。最重的夸克（顶夸克）的质量大约是电子质量的350000倍，也就是说，这种夸克与希格斯海相互作用的强度是电子与希格斯海相互作用强度的350000倍；在希格斯海中加速这样的夸克非常困难，这也就是这种夸克为什么那么重的原因。如果我们将粒子的质量比作一个人的名气，那希格斯海就可以算作是狗仔队：平常的老百姓可以随意地在成群的摄影师中穿行，但是政治名人或电影明星要想过去可就得费点劲了。[9]

我们这样来理解一个粒子的质量为什么不同于另一个粒子的质量当然可以，但是人们现今还不能从根本上解释已知粒子是怎样与希格斯海相互作用的。因而，人们也就无法从根本上解释基本粒子为什么具有实验上观测到的那个质量。不过，很多物理学家相信，要不是有希格斯海的存在，所有的基本粒子都应该像光子一样无质量。事实上，我们将会看到，这种情况很有可能就是宇宙最初时刻的情况。

冷却宇宙中的统一

水蒸气在100摄氏度的时候凝结成水，水在0摄氏度的时候结冰；理论研究告诉我们，希格斯场在千万亿（$10^{15}$）摄氏度的时候出现非零的真空期望值。这个温度大约是太阳中心温度的1亿倍；人们相信，这个温度也是大爆炸之后千亿分之一（$10^{-11}$）秒时的宇宙温度。在大爆炸之后$10^{-11}$秒之前，希格斯场剧烈振荡但平均值为零；就像100摄氏度以上的水在沸腾一样，这样温度的希格斯场也像沸腾了一

样难以平静。希格斯海瞬间蒸发。没有了希格斯海，各种粒子的加速运动不受任何阻碍（狗仔队消失了），也就是说，所有的粒子（电子、上夸克、下夸克和其他的各种粒子）的质量都一样——全都是零。

这样的认识部分解释了希格斯海的形成为什么要用宇宙相变来描述。在由水蒸气到水和由水到冰的相变过程中，有两件重要的事情发生。一个是外在性质上的改变，另一个是对称性的减少。在希格斯海的形成过程中，我们也可以看到这两方面的改变。一方面，有一个性质上的转变：曾经无质量的粒子突然获得质量——我们今天看到的粒子所具有的质量。另一方面，这一转变过程伴随着对称性的减少：希格斯海形成之前，所有的粒子具有相同的质量——零质量，事物具有高度的对称性。如果你互换两种不同粒子的质量，没有任何人会发现，因为所有粒子的质量都一样。但是希格斯场凝聚之后，粒子的质量变为非零——且不等——值，质量之间的对称性不复存在了。

事实上，由于希格斯海的形成而减少的对称性非常多。$10^{15}$摄氏度以上，希格斯场还没有凝聚。这个时候，并不只是所有的物质粒子无质量，由于没有希格斯海的阻碍，所有力的粒子也无质量（如今，弱核力的信使粒子——W粒子和Z粒子的质量分别是质子质量的86倍和97倍）。这时，正如20世纪60年代由谢尔顿·格拉肖、史蒂文·温伯格和阿卜杜斯·萨拉姆最先发现的那样，所有力的粒子的无质量性意味着另一种美妙的对称性。

早在19世纪后期，麦克斯韦就认识到，原本人们以为彼此截然不同的电和磁，实际上是同一种力——电磁力——的不同方面（参见本书第3章的内容）。麦克斯韦发现，电和磁彼此补足；两者是一种更具对称性的统一整体的阴阳两面。格拉肖、温伯格和萨拉姆发现了这一故事的下面几章。这几位科学家发现，在希格斯海形成之前，力的粒子并不仅是有相同的质量——都是零——光子、W粒子和Z粒子在所有其他方面本质上也都一样。[10]将一片雪花转过某些特殊角度，从

而使其不同的梢部交换位置，雪花看起来完全没发生变化；在没有希格斯海的情况下，互换电磁力与弱核力的信使粒子——也就是将光子与W粒子和Z粒子交换，物理过程也不会发生变化。雪花在旋转变换中保持不变意味着某种对称性（转动对称性）的存在；互换不同力的信使粒子后物理过程不受影响也意味着某种对称性，这种对称性由于技术上的原因被称为规范对称性，其意义极其深刻。这些粒子的作用是传递力——它们是不同力的信使粒子——因而，这些粒子之间的对称性也就意味着力之间的对称性。所以，温度足够高的时候，今天无处不在的希格斯海将会被蒸发掉，此时的弱核力与电磁力别无二致。也就是说，随着希格斯海在足够高的温度下蒸发掉，弱核力和电磁力之间的区别也随之蒸发。

格拉肖、温伯格与萨拉姆扩充了麦克斯韦一个世纪前的陈年发现，他们发现电磁力和弱核力不过是同一种力的两面。这3位物理学家成功将电磁力和弱核力统一为我们今天所称的电弱力。

现如今，电磁力和弱核力之间的对称性已经不再明显地表现出来。这是因为，随着宇宙冷却，希格斯海形成；光子、W粒子与Z粒子分别以不同的强度与凝聚的希格斯场相互作用。这就是问题的关键所在。光子可以在希格斯海中自由穿行——就像不再引起狗仔队兴趣的过气明星一样——仍然保持无质量。但是W粒子和Z粒子——就像比尔·克林顿和麦当娜一样——则会受到希格斯海的阻碍，它们分别获得了各自的质量，一个是质子质量的86倍，另一个是质子质量的97倍（注意：狗仔队这个比喻可与大小无关，但质量是与大小有关的）。于是，我们看到的电磁力和弱核力就显得如此不同。两者背后潜在的对称性“破缺”了，深埋在了希格斯海中。

这样的结论令人非常吃惊。在今天的温度下看起来完全不同的两种力——与光、电和磁这些现象有关的电磁力；以及与辐射衰变有关的弱核力——在本质上竟然是同一种力的两个不同方面，而之所以看起来不同的原因竟然是非零的希格斯场遮盖了这两种力之间的对称

性。这也就意味着，我们平常以为完全虚无的空间——真空，一无所有——竟然扮演了一个关键的角色，正是它使得世界看起来如我们所见。只有蒸发掉真空，将温度提到足够高从而使希格斯场被蒸发掉——也就是说，希格斯场在整个空间中的平均值为零——潜藏在大自然中的完整对称性才会显现出来。

格拉肖、温伯格与萨拉姆发展这些想法的时候，人们还没能在实验上发现W粒子和Z粒子。带着对理论的力量和对称性的美的坚定信念，这3位物理学家满怀信心地向前推进理论。他们的勇敢得到了回报。实验适时地发现了W粒子和Z粒子，从而确证了电弱理论的正确性。格拉肖、温伯格与萨拉姆透过肤浅的表面现象——他们克服了一无所有造成的障眼法——发现了将自然界中4种力中的两种联系起来的深妙对称性。由于成功统一了电磁力和弱核力，这3位物理学家于1979年被授予诺贝尔物理学奖。

大统一

我还是个大一学生的时候，会时不时地拜访我的导师，物理学家霍华德·乔奇。大部分时候我没什么话好说，不过那一般没什么关系。乔奇总有一些令人兴奋的东西与感兴趣的学生分享。有那么一次，乔奇特别来劲地讲了大约1个多小时，满黑板都是符号和方程。我也听得非常激动，止不住地点头。但是坦率地讲，我当时什么都没听懂。几年之后我才认识到，乔奇当时是在讲一些实验方案，他要用这些方案来检验他当时的理论发现——大统一理论。

电弱理论的成功使人们很自然地提出一个问题：如果自然界中的两种力在宇宙的早期曾是统一的整体的话，那么有没有可能在宇宙演化的更早期温度更高的时候，自然界中的3种力甚至全部4种力之间的差别全都消失掉，从而导致自然界有更高的对称性？这个问题实际上提出了一种激动人心的可能性：大自然中很可能只有一种基本的力，这种基本的力在一系列的宇宙相变之后分化成了我们今天看到的截然

不同的4种力。1974年，乔奇和格拉肖提出了第一个奔向这一完全统一目标的理论。他们的大统一理论，以及乔奇、海伦·奎因和温伯格后来的理念，向人们昭示了4种力中的3种——强核力、弱核力和电磁力——都是一种统一力的一部分；而这种统一力只在大爆炸之后的$10^{-35}$秒之内温度高达万亿亿亿（$10^{28}$）摄氏度——太阳中心温度的数十万亿亿倍——的极端条件下才能存在。前面提到的几位物理学家提出，当温度高于$10^{28}$摄氏度时，光子、强核力的胶子，以及W粒子和Z粒子全部可以彼此自由交换——一种比电弱理论的规范对称性更强的规范对称性——而不会带来任何可观测的物理后果。乔奇和格拉肖据此提出，在这样的高能高温条件下，3种非引力的力粒子之间具有完整的对称性，也就是说，3种非引力的力之间具有完整的对称性。[11]

乔奇和格拉肖的大统一理论还告诉我们，之所以我们今天不能看到这种对称性——将质子和中子紧紧胶连在原子核内部的胶子看起来与弱核力和电磁力完全不同——是因为在温度小于$10^{28}$摄氏度的时候，出现了另一种希格斯场。这种希格斯场被称为大统一希格斯场（为了避免混淆，我们可以将参与电弱统一的希格斯场称为电弱希格斯场）。类似于它的电弱表弟，大统一希格斯场在$10^{28}$摄氏度以上也会剧烈波动；计算又告诉我们，当宇宙温度低于$10^{28}$摄氏度时，大统一希格斯场就会凝聚成非零值。而且，同电弱希格斯场一样，大统一希格斯海形成的时候，宇宙也经历了一次对称性减少的相变过程。这时，由于大统一希格斯海对胶子的影响不同于其对其他粒子的影响，因而强核力与电弱力分开，于是之前唯一的一种非引力的力就一分为二成了两种。再过一点时间，温度又下降了很多之后，电弱希格斯场凝聚，弱核力与电磁力也彼此分开。

但是这样一个美妙的想法——大统一，还没有得到实验的确证（不同于电弱统一）。而且正相反的是，乔奇和格拉肖的原始理论预言了宇宙早期对称性在今天的一条遗迹，这就是质子衰变；大统一理论预言质子可以衰变成其他不同的粒子（比如说反电子和π子之类

的）。但是在几个精心打造的地下实验室中，历经数年的艰苦实验寻找——这种实验就是很多年前乔奇在他的办公室里兴奋地向我描述的那种实验——并未发现质子衰变，一无所获；这样的实验结果宣判了乔奇和格拉肖理论的命运。不过，在那以后，人们又提出了原始理论的各种变种，这些变种理论可以不被质子衰变实验排除；但是同样的，所有的这些变种理论都没有得到实验的任何支持。

多数物理学家认为大统一是粒子物理中的一个伟大想法，只是还没有真正实现。因为人们已经在电磁力和弱核力中证明了统一和宇宙相变的有效性，所以在很多人看来，将其他的力包括到统一的框架下只是时间问题。我们在第12章中将会看到，在这个方向上，人们已经利用另一种不同的方法向前迈了一大步，这种方法就是超弦理论。尽管这一理论还不完善，仍在发展；但是人们第一次将所有的力——包括引力在内，统一到了一个框架下。但是，即使只在电弱理论的框架下看，有一点也是非常清楚的：我们现在看到的宇宙所展现出来的并不只是早期宇宙的丰富对称性的残余。

以太的回归

对称性破缺的概念，及其通过电弱希格斯场的实现，在粒子物理和宇宙学中扮演着核心角色。不过我们的讨论可能会使你想到下面这个问题：如果希格斯海是一种看不见但却填满了我们过去认为虚无的真空的东西，那它岂不就是远古的以太概念的现代化身？我的回答是：既对也不对。让我稍作解释：说它对，是的，希格斯海的某些方面还真有点像以太。和以太一样，凝聚的希格斯场遍穿整个空间，无处不在，弥散于每一种物质中，而且作为真空的一个不可移除的性质（除非我们将宇宙加热到$10^{15}$摄氏度，但是这个我们做不到），希格斯海重新定义了一无所有这个概念。说它不对，因为希格斯海也有不同于原始的以太的地方。人们之所以引入以太的概念是为了替光波的传播找到介质，就像声波可以在空气中传播，那时的人们也要找到光波的传播介质。可希格斯海却与光的运动毫无关系，它对光速一点影

响都没有，因而20世纪初通过研究光的运动而排除了以太的实验，对希格斯海一点威胁都没有。

而且，由于希格斯海对任何匀速运动的东西都没有任何影响，它就不能像以太一样找出一种特别的观测点。正相反，即使有希格斯海存在，所有的匀速观测者彼此地位仍然相同，希格斯海并不与狭义相对论冲突。当然，这些观测事实不能证明希格斯海的存在；我们只能说，这些观测事实表明，尽管希格斯海的某些特征类似于以太，但是希格斯海并不与任何理论或实验冲突。

如果真有希格斯场所成的海，那么，它所带来的物理后果将在未来几年内在实验上得以确证。首先，就像电磁场是由光子组成的，希格斯场也是由某种粒子组成的，这种粒子的名字自然就是希格斯粒子。理论计算表明，如果真有弥漫于整个空间的希格斯海，那么希格斯粒子就应该能在大型强子对撞机（LHC）上的高能对撞残片中找到；LHC是一台巨大的原子对撞机，坐落于瑞士日内瓦的欧洲核子中心（CERN），将于2007年投入运行。简单地说，质子之间大量的正面对撞可能将一些希格斯粒子撞出希格斯海，这就好像水下的高能碰撞可能将某个$H_2O$分子敲出大西洋一样。实验将适时地告诉我们以太的这一现代版本究竟真的存在还是只能拥有以太以前被遗弃的命运。这一问题非常重要，因为我们已经看到，凝聚的希格斯场在当前的基本物理理论框架下扮演着深远而关键的角色。

如果我们不能发现希格斯海，那么我们就得好好反思一下我们已经用了30年的理论体系了。但是一旦找到了希格斯海，那就是理论物理学的一次重大胜利：它将使人们再一次见识对称性的威力，在我们闯入未知领域的时候，正是对称性帮助我们正确地进行数学推演。除此之外，确认希格斯海的存在还有其他的意义。首先，这将会直接证实今天各不相同的力在远古时期曾是统一的整体。其次，我们将会认识到我们长久以来对真空的直观认识——将一个区域内的一切全部取出从而使得该区域的能量和温度降到有多低就多低的水平——非常的

幼稚。最空的真空也不必非得是空无一物。因此，不必使用精神方面的概念，就在我们寻找关于空间和时间的答案时，我们也能碰到亨利·摩尔的想法（见第2章）。对于摩尔来说，普通的真空概念毫无意义，因为空间总是盛装着神圣的精神。对于我们来说，普通的真空概念同样难以琢磨，因为真空中可能总是盛装了希格斯海。

## 熵与时间

图9.2中的是我们所讨论过的相变时间线，这幅图可以帮助我们更好地掌握从大爆炸开始的那一刻到你厨房中的鸡蛋煮熟时宇宙中发生的一系列重大事件。当然，关键的信息仍然隐藏在令人眼花缭乱的表面之下。但是你一定要记住，知道事物如何开始——《战争与和平》的书页顺序，可乐瓶中的二氧化碳分子，大爆炸时的宇宙状态——是知道事物如何演化的关键。只要还有余地，熵就会增加。只要开始很低，熵就会增加。如果《战争与和平》的页序本来就是乱的，再怎么抛洒它，它还是只能保持打乱的页序。如果宇宙开始于一个完全无序的高熵状态，那么以后的宇宙演化也只能保持这一无序状态。

图9.2所示的历史明显不是连续的、不变的无序度的编年史。即使某些对称性因宇宙相变而消失，宇宙整体的熵还是在稳固的增加中。但是，在最开始宇宙必然处于高度无序的状态。这一事实允许我们将时间上的“向前”与熵增的方向联系起来，但是我们仍需为刚刚诞生时的宇宙那不可思议的低熵——不可思议的高度均匀性——找到一个解释。这就要求我们继续回溯并去努力理解最初时刻——图9.2中模糊地带——究竟是怎样的。现在我们就开始这个任务。

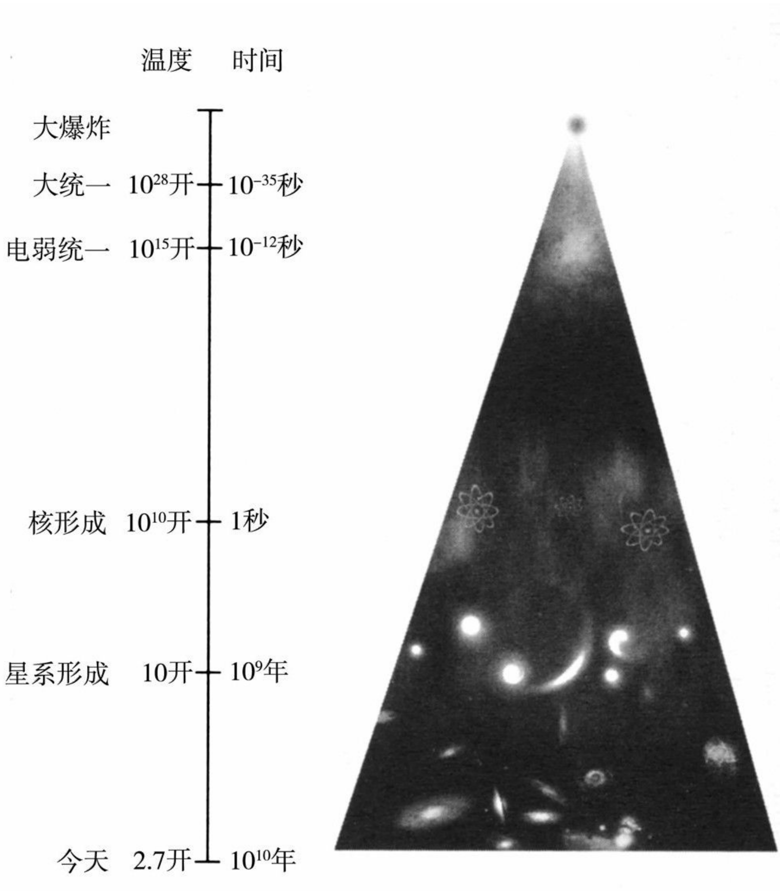



图9.2　图释宇宙学的标准大爆炸模型的时间线。

# 第10章　解构大爆炸

## 到底是什么爆炸了

人们通常会认为大爆炸是关于宇宙起源的理论，这是一种错误的认识，大爆炸并不是这样的理论。在前面两章中，我们曾讨论过大爆炸理论的部分内容。大爆炸理论描述的是宇宙在某一瞬间突然诞生之后的演化情况，至于在那个瞬间到底发生了什么，不是大爆炸理论能

够回答得了的问题。大爆炸理论只说假定最开始的时候有一场爆炸；然后它就把爆炸扔到一边了，并没有认真地讨论爆炸的问题。大爆炸理论并没有告诉我们到底是什么爆炸，为什么爆炸，怎样爆炸，而且老实地讲，大爆炸理论都没有告诉我们究竟是否真的有这样一场爆炸。[1]事实上，只要你稍稍动动脑筋，你就会认识到大爆炸理论给我们出了个大难题。在宇宙的最初时刻，物质和能量处于极高密度的状态，这时引力46远大于各种力。但我们知道，引力是一种吸引力。引力会使事物彼此靠近。这样的话，又是哪种向外的力驱动空间扩展呢？看来，在大爆炸的时刻，一定有某种强大的排斥力扮演了关键的角色。但自然界中的哪一种力可以担当这个角色呢？

这个宇宙学中最基本的问题几十年都没有解决。然后，时间到了20世纪80年代，爱因斯坦的一个老想法以一种新颖的形式得以复兴，摇身一变成了暴涨宇宙学。靠着暴涨宇宙学，人们终于为大爆炸找到了它该有的力——引力。这一答案很令人吃惊。物理学家们认识到，在适当的情形下，引力可以是排斥力；而根据暴涨理论，在宇宙的早期，正好有引力为排斥力所需要的必要条件。在一段就连纳秒都可算是永恒的极短时间内，早期宇宙为引力提供了一个舞台；在这个舞台上，引力可以表现出它不为人所知的一面——排斥性，这个时期的引力毫不留情地驱赶着空间中的一切，使之彼此远离。早期宇宙中，引力的排斥性如此之强，远远超出了人们以前想象的大爆炸强度。在暴涨宇宙学的框架下，早期宇宙的膨胀因子大到令人难以想象，远不是标准的大爆炸理论所预言的那样。暴涨宇宙学极大地扩充了宇宙学的内涵，与之相比，我们过去对宇宙学的认识渺小得简直就像千亿星系中的一颗星星。[2]

我们将在本章以及下面几章中讨论暴涨宇宙学。我们将会看到，暴涨宇宙学可以作为标准大爆炸宇宙学的“前端”，它对宇宙最初时刻的解释与标准的大爆炸宇宙学的解释有天壤之别。暴涨宇宙学解决了标准大爆炸宇宙学力所不及的问题，并且给出了一系列的预言；这些

预言有的已经被实验验证，有的将在未来的几年内接受实验的检验。而最令人惊奇的是，暴涨宇宙学向我们展示了量子过程是如何通过宇宙膨胀将细小的波纹印到空间的结构中去，从而在夜晚的天空留下可见的烙印。除了这些成就，暴涨宇宙学还能帮助我们更加深刻地理解早期宇宙是如何获得它那极低的熵，而这将使我们比以往任何时候都更加接近时间之箭的解释。

爱因斯坦与排斥性的引力

1915年，在为他的广义相对论写下最后一笔后，爱因斯坦立即将这一理论应用到多种问题中。其中之一就是长期困扰着牛顿方程的水星近日点的进动问题。实验观测告诉我们，水星并不是按照同样的轨道绕着太阳运动，它的运行轨道总要比前一次的运行轨道改变一点。牛顿理论对于这一现象的理论计算值与实验观测值之间有一个小小的差别。爱因斯坦用他的广义相对论计算了这个问题，结果他算出来的值与实验观测值精确相符，这样令人吃惊的结果强烈地刺激了爱因斯坦的心脏。[3]爱因斯坦又把他的广义相对论应用到另一个问题上：远方恒星所发出来的光在经由太阳到达地球的时候，其轨迹会因为时空弯曲而弯曲，爱因斯坦计算了这个弯曲量。1919年，两队天文学家——一队驻扎在非洲西海岸的普林西比岛，另一队在巴西——检验了这一预言。在有日食存在的情况下，天文学家们观测了刚好擦着太阳表面而来的星光（擦着太阳表面而来的星光受太阳的影响最厉害，而只有在日食的时候才能观测到这些星光），并将这次观测的结果与地球处在太阳和远方恒星中间时（这个时候，远方星光到达地球的轨迹几乎不会受到太阳引力的影响）所拍摄的星光照片相比，从比较中得到的弯曲角再次证实了爱因斯坦的理论计算。当观测结果发表出来的时候，爱因斯坦一夜成名。坦率地讲，当时的爱因斯坦可谓意气风发。

但是除了最初的成功，爱因斯坦许多年都不愿接受广义相对论方程带来的数学结果，因为他不满意将广义相对论应用到最浩渺的理论

挑战——理解整个宇宙——时所得到的结果。早在第8章讨论过的弗里德曼和勒梅特的工作之前，爱因斯坦就认识到，广义相对论的方程意味着宇宙不可能是静态的；空间的结构可以拉伸也可以收缩，就是不能保持固定的大小。这也就意味着宇宙可能有一个明确的开端，而当宇宙被极大的压缩时，宇宙可能也会有一个明确的尽头。爱因斯坦坚决抵制广义相对论带来的这一结果，因为他和其他人都“知道”宇宙是永恒的，在大尺度上，宇宙恒久不变。因而，尽管广义相对论拥有美和成功，爱因斯坦还是再次打开了他的笔记本，试图修改方程，使其能够满足他们那个时代的人对于永恒宇宙的认识。这并没花他多少时间。1917年的时候，爱因斯坦就实现了这一目标，而他的方法就是为广义相对论方程引入新的一项：宇宙常数。[4]

爱因斯坦引入这样一项的办法不难掌握。任意两个物体，不管是棒球、行星、恒星、彗星或是你有的什么东西，它们之间的万有引力都是吸引力，因而万有引力总在试图拉近两个物体。地球和一个向上跳跃的舞蹈演员之间的万有引力会使舞蹈演员减速，在达到最大高度后落回地面。如果舞蹈指导者要求舞蹈演员浮在空中摆一个静态造型，那么，在舞蹈演员和地球之间就要有一种排斥力来平衡地球和舞蹈演员之间的万有引力：只有在万有引力与这种排斥力精确相消的条件下舞蹈演员才能摆出静态造型。爱因斯坦认识到，对于整个宇宙来说，这一推论也成立。引力会减慢舞蹈演员上跃的速度，还能减慢空间的膨胀。没有能平衡万有引力的排斥力的话，舞蹈演员摆不出静态造型——她不能浮在固定的高度上；同样的，没有起平衡作用的排斥力的话，空间也不可能静止——空间不能保持固定的大小。爱因斯坦之所以要引入引力常数，就是因为他发现方程中有了这样一项后，万有引力就可以提供所需的排斥力。

但这样的数学项所暗含的物理意义究竟是什么呢？宇宙常数到底是什么，由什么组成，又是怎样抵消平常的吸引性万有引力从而产生向外排斥的力呢？对爱因斯坦工作——可追溯到勒梅特——的现代诠

释告诉我们，引力常数是一种奇怪的能量形式，均匀地填充于整个空间。而我之所以说它“奇怪”，是因为爱因斯坦的分析并没有告诉我们这种能量来自哪里；而且，我们马上还会看到，他所采用的数学描述方式又保证了这些能量不会来自我们熟悉的质子、中子、电子或光子之类。今天的物理学家们在讨论爱因斯坦的宇宙常数时会使用“空间本身的能量”或“暗能量”这样的术语；这是因为，如果真有一个宇宙常数的话，空间就会充满你不能直接看到的某种透明的、无形的存在；被宇宙常数占据的空间仍然是黑暗的（这样的说法有点像很老的以太概念和年轻点的非零希格斯场概念。而后者和宇宙常数之间倒真的不只仅仅相似那么简单，它和宇宙常数的确有某种联系，我们马上就会看到这一点）。不过即使说不清宇宙常数的起源或身份，爱因斯坦也能算出它对于引力的意义；而爱因斯坦的结果，的确有不同寻常之处。

要想弄明白这点，你必须知道一个我们还没讲过的广义相对论性质。在牛顿的引力理论中，两个物体之间的引力强度只取决于两件事：物体的质量及其间的距离。物体的质量越大，彼此靠得越近，将物体拉近的万有引力也就越强。在广义相对论中情况也差不多是这样，只不过在爱因斯坦的方程看来，牛顿的注意力不应该只集中在质量上。广义相对论告诉我们，除了物体的质量（当然还有距离）之外，能量和压强也对引力场的强度有贡献。这一点很关键，我们现在来看看这究竟意味着什么。

假定现在是25世纪，而你被关在智力大厦内。最新的惩教实验正在进行，这一实验试图通过精英模式来训诫白领犯人。实验中，每个罪犯将获得一道题目，只有解决了这道题目，罪犯才能重获人身自由。住在你旁边囚室的犯人得到的题目是：为什么《盖里甘的岛》[47]的重新上映会在22世纪掀起狂潮，从而一举成为有史以来最受欢迎的电影？这个问题显然并不简单，看来这位可怜的人有得忙了。你抽到的问题要稍稍简单一些。有两个完全一样的固态金块，大小一样，金

的纯度也完全一样。你所面临的挑战是：找出一种办法，使两块金块分别被轻轻地放到固定的精准天平上时，天平的示数不一样；前提是你不能改变其中任何一块中的物质数目，也就是说你不能采用切削焊割这样一些办法。如果你把这个问题交给牛顿，他马上就会告诉你这个问题无解。因为根据牛顿的理论，等量的金意味着等大的质量。既然测量金块所用的天平完全一样，刻度固定，施加于其上的地球引力也会一致。所以，牛顿自然会说，这样的两块金块一定会有相同的示数，没有任何例外的可能性。

但是作为一个25世纪的人，你在高中阶段就学过广义相对论了，所以你能够解决这道题目。广义相对论说两个物体之间的引力大小并不仅仅取决于物体的质量[5]（当然还有两者之间的距离），还取决于那些可以加到每个物体总能量上的额外贡献。所以我们还可以在物体的温度上做点手脚。温度是组成金块的金原子运动的平均快慢的量度，也就是说，温度是金原子活性的量度，反映的是金原子的平均动能大小。于是你就可以知道，只要加热金块，金原子就会运动得更猛烈，因而被加热的金块就会比没有被加热的金块更重。牛顿并不知道这个事实[1磅（1磅≈0.4526千克）重的金块温度升高10摄氏度，重量增加千万亿分之一，其效应非常之小]，但是你知道，所以你就能够逃脱囚困。

你回答了这个问题，但是你的罪很重，假释官在最后一刻决定你还必须再受一次考验。这次他们给了你两个完全一样的玩具，就是那种打开盒子就有小人弹出的玩具。现在你要做的是再使这两个玩具有不同的重量。这次对你的限制苛刻了一点，你不但不被允许改变每一个玩具的质量，还不可以改变它们的温度。要是把这个问题给牛顿，他就只好一直被关在大厦里了。因为他还是只能给出那套解释：物体质量一样，因而重量必定一样，所以本题无解。而你呢？你还是可以求助于广义相对论：你可以把一个玩具中的小人紧紧地压在盒子底，让另一个玩具中的小人处在弹出的状态。为什么要这么干？答案是压

缩中的弹簧比正常伸展的弹簧具有更多的能量。你花了力气压缩弹簧相当于给了弹簧能量，弹簧的弹力反映的就是你的劳动成果，使小人向外弹出的就是这股弹力。于是，我们又可以回到之前的说法：爱因斯坦告诉我们，任何多余的能量都会影响引力，从而使物体具有额外的重量。因此，小人被紧紧压到盒子底部的玩具将比小人正常伸展的玩具重那么一点点。牛顿不了解这一点，但是你知道，所以你就一定可以重获自由身。

第二个问题的解决方案向我们透露的正是广义相对论的微妙之处。爱因斯坦在他那篇有关广义相对论的论文中，用数学为我们展示了万有引力并不仅仅取决于物体的质量，也不仅仅取决于能量（比如热这种能量），还取决于可能具有的压强。要理解宇宙常数的问题，我们就非得了解广义相对论的这一特点。而这也就是为什么向外的压强——比如压缩的弹簧所具有的——被称为正压的原因。显然，正压将对万有引力有正的贡献。于是，关键之处来了：压强，并不同于质量和总能量，在某些情况下，某些区域的压强可以为负，这样的压强有向里吸而不是向外推的效应。虽然乍听之下没那么奇怪，但以广义相对论的角度来看，负压会导致某些非常奇怪的事情：正压可以对普通的万有引力有贡献；负压贡献的却是“负”引力，也就是说，负压贡献的是排斥性的万有引力！[6]

爱因斯坦的广义相对论带来的这令人错愕的结果，打破了人们200年来的固有信念——万有引力只能是一种吸引力。行星、恒星和星系，的确如牛顿告诉我们的那样，一直展示出来的是吸引力。但是，在某些情况下，当压强变得非常重要（在我们日常生活的条件下，压强带来的引力贡献完全可以忽略），特别是负压（对于普通物质，如质子电子之类，压强是正的，因而宇宙常数不可能来自人们熟悉的普通物质）变得非常重要的时候，万有引力的效应可能令牛顿目瞪口呆，因为万有引力可能是排斥力。这个结果对于由此而来的诸多推论非常重要，而且很容易被错误的理解，所以我要强调一下它的关

键之处。在广义相对论的框架下，引力和压强既有联系又有区别。压强，或者更准确地说压强差，可以以自己的方式，一种非引力的力的面目出现。潜水的时候，你的耳膜就会感觉到压强差，因为耳膜外水压和耳膜内的气压彼此不同。这种说法完全没有问题。但在我们讨论的压强和引力问题中的压强则完全不同。根据广义相对论，压强对引力场有贡献，因而可以间接显示出力的效应——通过对万有引力有贡献体现出来。压强，虽然不同于质量和能量，但也是引力的一个来源。特别要注意的是，要是某一区域内的压强为负，那它在这个区域内就会贡献出排斥性的万有引力，而不是吸引性的万有引力。

这就意味着，假如压强为负，那么来自于质量和能量的普通吸引性的万有引力，和来自于负压强的排斥性的万有引力之间会存在一种抵消。如果区域内的负压强大到一定程度，那么排斥性的万有引力就会起主导作用；万有引力就会把物质彼此拉近而不是推开。于是宇宙常数就可以登场了。爱因斯坦加到广义相对论方程中的宇宙常数项相当于在空间中均匀地布满能量，方程又告诉我们这种能量具有负压强。而且，来自于宇宙常数负压的排斥性万有引力会超越来自于正能量的吸引性万有引力，因而排斥性的万有引力将起主导作用：宇宙常数表现出的是排斥性的万有引力。[7]

在爱因斯坦看来，这简直就是对症下药。遍布于整个宇宙的普通物质和辐射会展现出吸引性的万有引力，从而会将每一块空间拉向彼此。而新的宇宙常数项，同样遍布于整个宇宙，则会展现出排斥性的万有引力，从而将每一空间区域推离彼此。准确地调节该项的大小，爱因斯坦就可以用新的排斥性万有引力来平衡原有的吸引性万有引力，从而获得静态宇宙。

爱因斯坦还发现，由于排斥性的万有引力来自于空间本身的能量和压强，因而它的强度具有累加性。也就是说，空间间隔越大，这种排斥性的万有引力也就会越大，这是因为更大的空间意味着更多的外推力。而在地球或者太阳系的尺度上，这种排斥性的万有引力微乎其

微。只有在跨度巨大的宇宙尺度上，这种力才会变得明显。这样一来，在小到我们日常生活的尺度上，牛顿理论的成功和爱因斯坦自己的引力理论就不会有任何矛盾之处。总而言之，爱因斯坦可以舒舒服服的享受一下了：他既拥有了已由实验确认的广义相对论性质，又得到了一个静态的宇宙，这样的宇宙既不会膨胀也不会收缩。

有了这样的结果，爱因斯坦无疑会长舒一口气。如果10年苦心钻研所得到的广义相对论方程竟然不能同人们每天晚上仰望星空就能看到的静态宇宙事实相符的话，那爱因斯坦该有多揪心。但是，正如我们所见，20年之后情况急剧变化。1929年，哈勃发现人们对天空的简单认识可能是错误的。系统的观测告诉哈勃，宇宙可能并非处于静态，而是处于膨胀中。要是爱因斯坦相信原始的广义相对论方程的话，他本该在10多年之前就预言宇宙正在膨胀，而不是等到实验来发现这一事实。这样的实验发现毫无疑问可算作有史以来最伟大的发现之一——甚至可算作最伟大的发现。知道了哈勃的发现之后，爱因斯坦懊恼不已，他小心地将宇宙常数项从他的广义相对论方程中擦去了。爱因斯坦希望大家忘了宇宙常数的事，而大家的确也把宇宙常数忘记了，这一忘就是几十年。

但时间到了20世纪80年代，宇宙常数死灰复燃，并且以宇宙学思想史上最炫目的形式出现在大家面前。

蹦跳的青蛙和过冷却

对于向上飞去的棒球，你可以用牛顿方程（或者更准确的爱因斯坦广义相对论方程）来计算出它接下来的运动轨迹。一旦你算过一次，你就会彻底搞清楚球的运动。但是，这里还有个问题：最初是谁或者说是什么使球向上飞了出去？虽然你可以用数学算出球后来的运动情况，但球最开始的时候是如何启动的呢？对于现在的这个例子，答案当然很简单，就是棒球手扔出去的呗（当然，它也有可能是从停靠在一边的奔驰车的挡风玻璃上弹飞的，不管怎样，你总是知道

的）。但所考虑的若是一个更复杂的类似问题，比如说宇宙膨胀的广义相对论解释，答案可能就不是这么明显了。

广义相对论方程允许宇宙膨胀。这一点最先由爱因斯坦和荷兰物理学家威廉·德·西特证明，随后弗里德曼和勒梅特也发现了相同的结论。但是，正如牛顿方程不会告诉我们最开始是什么使棒球飞起来，爱因斯坦的方程也没有告诉我们宇宙是怎么开始膨胀的。许多年来，宇宙学家们一直只把空间开始向外膨胀当成解释不了的初始条件，然后用方程来研究之后发生的事情。现在你明白我为什么会在前面说大爆炸理论没有告诉我们哪怕是一点有关爆炸本身的事情了吧？

1979年12月的一个晚上，一切都改变了。艾伦·古斯，斯坦福直线加速中心的一位年轻博士后研究员（现在是麻省理工学院的一名教授），告诉人们还有很多事情可以理解。是的，的确很多。尽管在20年后的今天，仍有很多细节未能搞清，但是艾伦·古斯的确做出了使宇宙学改头换面的重大发现，他的发现为大爆炸理论找到了一场爆炸，一场比所有人预想的都要大得多的爆炸。

古斯本不是一名宇宙学家，他原来的专业在粒子物理学领域。20世纪70年代末，古斯跟着康奈尔大学的亨利·泰研究大统一理论中希格斯场的有关问题。还记得我们在上一章中讨论过的内容吗？当希格斯场的值固定为某个特殊的非零值（这个非零值取决于希格斯场势能碗的具体形状）时，对称性自发破缺，希格斯场对空间区域贡献最低可能的能量。在早期宇宙中，温度极高的时候，希格斯场会强烈的波动，就像热铁碗中上串下跳的青蛙一样；但随着宇宙的冷却，希格斯场落到某个具体的值，这个时候其能量处于最低状态。

古斯和泰试图弄清的是：为什么希格斯场会延迟达到其最低能量态[图9.1（c）中势能碗的谷底]？如果我们用青蛙来类比古斯和泰的问题，那就是：如果在碗刚开始冷却的时候，青蛙的某一跳恰好跳到了碗的中央高地，那会怎样呢？而且，要是碗继续冷却，但是青蛙却

一直待在碗的中央（美美地享受那些蚯蚓），而不是滑到碗的底部，又会怎样呢？或者用物理学的术语来说，在宇宙冷却的时候，如果希格斯场的值恰好波动到势能碗的中央并且一直保持在那里会怎样呢？如果是这种情况，物理学家就会说希格斯场处于超冷状态；这一名词传达的意思是：即使宇宙的温度低于希格斯场处于最低能量态时的温度，希格斯场仍处于较高能量态（在日常生活中也有这样的现象，比如高度纯净的水，就有可能在温度低于零摄氏度——正常情况下的冰点——的时候仍未结冰。这是因为晶体需要依托杂质才能生长；没有杂质的话，晶体无法生长，液态水就无法结冰）。

古斯和泰之所以对这种可能性感兴趣是因为这种可能性可能与研究人员在各种大统一的尝试中所遇到的一个问题有关（这个问题就是磁单极问题[8]）。不过古斯他们还想到了另外的可能性，而这另外的可能性才是使其工作变得重要的真正原因。他们怀疑，伴随超冷希格斯场而来的能量——还记得吗？场的高度代表其能量，所以，只有当场的值在势能碗的谷底时场才会具有零能量——可能对宇宙的膨胀有重要影响。1979年12月上旬，古斯仔细地考虑了这些想法，得到了下面的结论。

达到稳定状态的希格斯场会在空间中填充能量，但并不仅仅如此。更为重要的是，古斯认识到，希格斯场还会对空间贡献均匀的负压。事实上，古斯发现，一旦将能量和压强考虑进来，达到稳定状态的希格斯场将具有和宇宙常数相同的性质：希格斯场在空间中填充能量和负压，这正是宇宙常数所具有的性质。于是，古斯就发现了超冷的希格斯场的确对空间膨胀有重要的影响：就像宇宙常数那样，希格斯场也能显现出使空间膨胀的排斥性引力。[9]

到了这里，因为你已经熟悉了负压与排斥性万有引力，所以你就可能会这样想：没错，古斯的确发现了一种来实现爱因斯坦宇宙常数想法的办法，但是那又能怎样呢？这有什么大不了的呢？宇宙常数的概念不是早就被扔到废纸篓里了吗？宇宙常数除了令爱因斯坦蒙羞外

还有什么用呢？重新发现已经60多年没人理的东西有什么可值得激动的呢？

暴涨

我们现在就来说说古斯的发现妙处在哪？古斯发现，虽然超冷希格斯场具有很多宇宙常数所具有的性质，但是它并不只有宇宙常数才有的性质。事实上，超冷希格斯场与宇宙常数有两点关键的区别——而正是这两点区别使得两者完全不同。

首先，尽管宇宙常数是一个常数——常数就意味着不随时间改变，因而宇宙常数贡献的是稳定不变的外推力——而超冷希格斯场并不是常数。我们来想想居于图10.1（a）的碗中高处的青蛙。它虽然会在那里待上一会，但迟早会随随便便的跳一跳——这倒不是因为碗太烫，而完全是因为青蛙总会待得有些无聊——而这会使青蛙不小心掉下去，然后就像图10.1（b）那样滑落到碗底。希格斯场的行为就类似于此。当温度变得太低出现剧烈振荡时，希格斯场在整个空间中的值可能会固定在其势能碗的中心高地处。但是量子过程会带来很多随机波动，这些随机过程会使希格斯场的值涨落，离开中心的位置，从而使得其能量和压强变为零。[10]古斯的计算表明，这种涨落会由于势能碗的具体形状不同而有所不同，在有些情况下会在极短的时间内发生，这一极短的时间可能会短到0.00000000000000000000000000000000001（$10^{-35}$）秒。随后，当时在莫斯科列别德夫物理研究所工作的安德烈·林德，和当时在宾夕法尼亚大学的保罗·斯坦哈特及其学生安德里亚斯·阿尔布莱奇发现了一种可以使希格斯的能量和压强在整个空间更有效也更均匀地变为零的办法（并且同时解决了古斯原始理论中的一些技术问题[11]）。这3位物理学家证明，如果势能碗像图10.2所示的那样，更加光滑更有坡度的话，有没有量子过程就不重要了：希格斯场的值会很快滑到谷底，就像一个从山顶落下的球那样。这些分析的结论就是：如果希格斯场非得像宇宙常数一样，它也只能在一个瞬间内像宇宙常数。

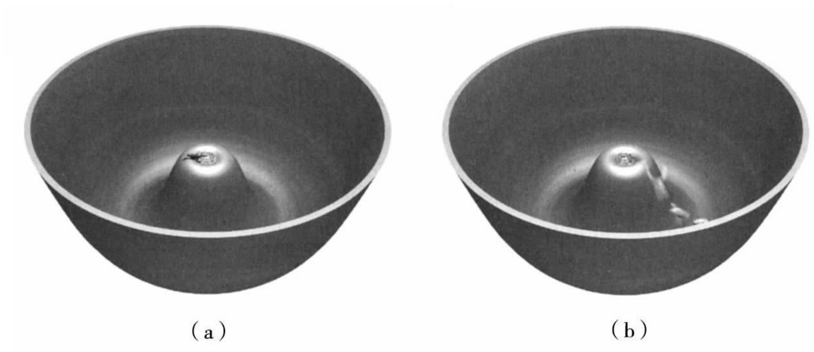


图10.1 （a）：超冷希格斯场陷在势能碗的高能位置上，就像碗中凸起处的青蛙一样。（b）：一般来说，超冷希格斯场很快就会跌落到最低能量态，就像跳回到碗底的青蛙一样。

第二个区别在于，爱因斯坦要仔细但却任意地选择宇宙常数的值——可以贡献到每一寸空间的能量和负压——以使向外的排斥力与来自宇宙间普通物质和辐射的向内吸引力精确的平衡；而古斯却能计算出他和泰所研究的希格斯场所贡献的能量和负压。古斯计算出来的值是爱因斯坦所选取值的大约1000000000000000000000000000000000000000000000000000000000000000倍。显然，这个数字非常巨大。所以，与爱因斯坦的宇宙常数所带来的外推力比起来，由希格斯场的排斥性引力所导致的外推力简直就是无限。



图10.2　更光滑更有坡度的凸起使得希格斯场更加容易滑到零能量谷，并且在整个空间中也更加均匀。

我们再来看看这两个情况——即希格斯场停留在势能碗中心的稳定点，处于高能负压态，这样的状态仅能维持极短的瞬间；但是只要希格斯场处于这样的态，就会产生极其巨大的外推力——合在一起会有什么结果。古斯发现，这两种情况合在一起意味着一场时间极短但却影响巨大的大爆发。换句话说，这就意味着我们得到了大爆炸理论没给我们的东西：一场爆炸，而且是一场规模巨大的爆炸。这就是古斯理论激动人心之处。[12]

古斯的突破带给我们的是如下的宇宙图景。很久很久以前，宇宙极端致密，全部的能量都由希格斯场携带，而希格斯场处于远离其势能碗最低处的某个值上。为区别于其他的希格斯场（比如为普通粒子生成质量的电弱希格斯场，又或者大统一理论中的希格斯场[13]），我们将这个希格斯场称为暴涨子场。48因为其负压的缘故，暴涨子场会生成巨大的排斥性万有引力，这股巨大的力量使得每一片空间区域远离彼此；用古斯的话来说，暴涨子使宇宙暴涨。这股巨大的排斥性万有引力大约存在了$10^{-35}$秒，虽然时间短暂，但是由于它实在太过巨大了，因而使得宇宙一下子膨胀到难以想象的地步。根据暴涨子场势能

碗的具体形状不同，宇宙可能轻易地膨胀$10^{30}$倍、$10^{50}$倍、$10^{100}$倍，甚至更多倍。

这样的数字令人非常惊讶。膨胀$10^{30}$倍——最保守的估计——到底是怎样一种概念呢？想象一下吧，这就好比在眨一下眼的千亿亿亿分之一的时间内，一个DNA分子膨胀到了银河系那么大。对比来看，即使这最保守的估计也比标准大爆炸模型在相同时间内的膨胀量大百亿亿倍，这一瞬间的膨胀甚至比标准大爆炸理论中140亿年的累计膨胀量还要大！在很多暴涨理论模型中，实际计算出来的膨胀倍数都要大于$10^{30}$，这样的膨胀导致宇宙空间异常巨大，即使用我们最先进的望远镜，也只能看到全部宇宙的一小点。根据这些模型的理论预言，宇宙中大部分区域所发出来的光还从未曾到达过太阳系；而且，即使到了太阳和地球都消亡的那一天，很多地方发出来的光还不会到达。如果整个宇宙有地球那么大的话，我们能够看到的范围就只有一粒沙那么大。

大概在大爆炸开始之后的$10^{-35}$秒，暴涨子场从中心的稳定点跌落下来，其在整个空间中的值滑落到势能碗的底部，排斥性的万有引力消失了。随着暴涨子场的值滑落下来，它所积攒的能量也将释放出来，而均匀地填充于膨胀中的空间的普通物质和辐射就是产生于这股能量，就好像清晨的薄雾落到小草上成了露珠一样。[14]这个时刻以后的故事就是标准的大爆炸理论了：大爆炸之后空间继续膨胀并开始冷却，物质粒子聚成了星系、恒星和行星之类的结构，这些天体慢慢演变成我们今天看到的样子，如图10.3所示。

古斯的发现——所谓的暴涨宇宙学——以及后来林德、阿尔布莱奇和斯坦哈特等人的改进版本，第一次解释了是什么使得空间膨胀。居于零能量之上的希格斯场能够提供一种使空间膨胀的外推力。古斯为大爆炸理论找到了一场爆炸。

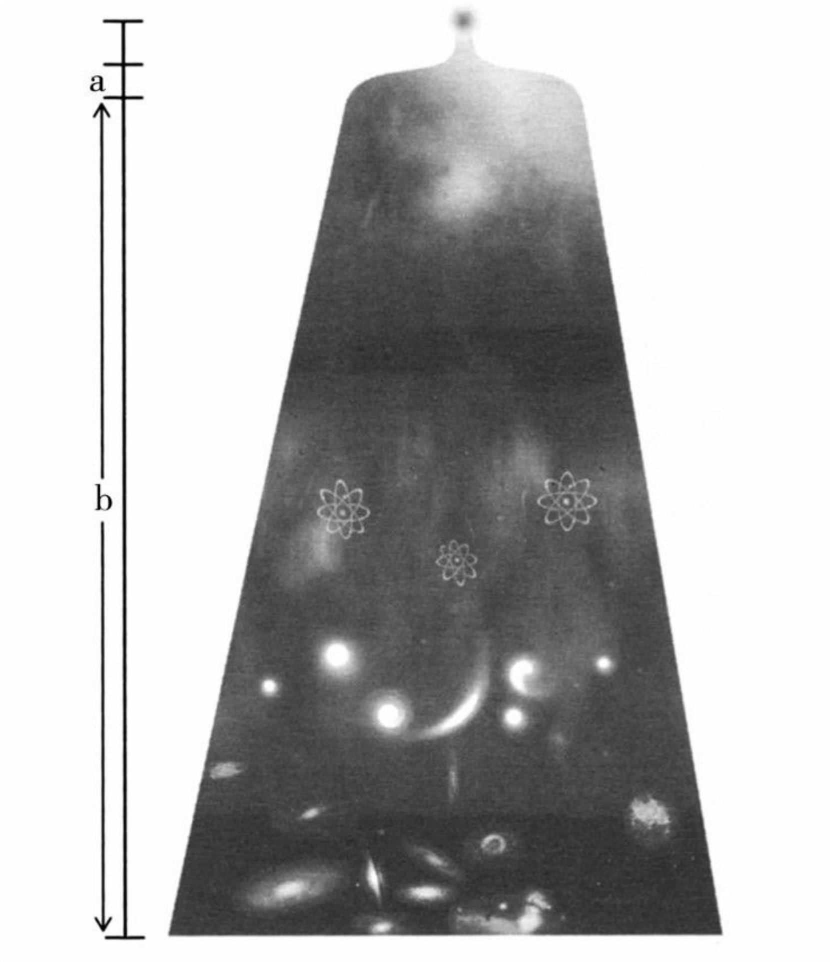



图10.3　（a）：暴涨宇宙学在宇宙演化的早期插入了一段迅速、猛烈爆发的空间膨胀时期。
（b）：猛烈的爆发之后，宇宙进入了大爆炸模型所提出的标准演化时期。

暴涨理论体系

古斯的发现很快就被人们视作一项重大进展，迅速成为宇宙学领域的主要研究对象。但是有两点值得我们注意。首先，标准的大爆炸模型假定爆炸发生在零时刻——宇宙开始的时刻，因而爆炸是一次创

生事件。暴涨理论中的爆炸发生在条件齐备的情况下——必须存在暴涨子场，这样排斥性万有引力的向外大爆发才能有原料——并且不必去符合宇宙的“创生”。这也就是说，我们最好将暴涨大爆炸想成是已经存在的宇宙所经历的一次事件，而不必将其视为就是创造宇宙的事件。将图9.2中的某些部分变得模糊，我们就得到了图10.3，这样的做法是要告诉大家我们对宇宙的起源仍一无所知。特别是，如果暴涨宇宙学真的是正确的，那我们就要问这些问题：为什么会有一个暴涨子场？而它的势能碗又为什么恰恰拥有暴涨能够发生的特别形状？为什么会存在空间和时间这样的事物使得这些讨论能够发生？还有，用莱布尼茨的话来问就是，为什么是有而不是无？

另一点值得我们注意的是，暴涨宇宙学并不是一个独一无二的理论。它是一个建立在万有引力可以是推动空间膨胀的排斥力这一认识基础之上的宇宙学理论体系。关于外推力的种种细节——一旦暴涨发生，它要持续多久，外推力的强度，宇宙在暴涨期间膨胀的倍数，暴涨结束的时候究竟有多少能量变成了物质，诸如此类的问题——主要取决于暴涨子场势能的尺寸和形状，而这一点又是我们的现有理论不能回答的问题。所以，很多年来，物理学家们研究了暴涨理论的各种可能性——势能的各种可能形状，以及可能的暴涨子场数目，等等——从中挑选出了那些能够与当前的天文学观测符合的理论。我们要知道的是，暴涨理论的某些方面并不是细节之所在，而是对于各种暴涨模型来说都很普通的东西。按照定义，爆炸本身就是这样的一种性质，任何一种暴涨模型都有一场爆炸。但是，所有的暴涨理论都有的另外一些先天性质则很关键，是我们要详细考虑的部分，因为这些性质可以解决一些原始的大爆炸宇宙学所不能解决的问题。

暴涨与视界疑难

这些问题中的一个就是所谓的视界疑难，它与我们之前提过的微波背景辐射的均匀性有关。回想一下，我们曾经讲过来自空间某一方向的微波辐射与来自空间任意其他方向的微波辐射在极高的精确度

（千分之一度的水平）上彼此相符。这一观测事实非常重要，因为它证明了空间的各向同性，可以大大简化宇宙学模型。在前面的章节中，我们曾经利用这种各向同性神奇的限制了空间可能的形状并探讨了均一宇宙时间。视界疑难这一问题在我们试图解释宇宙如何拥有均一性时出现。宇宙中的广大区域究竟是如何拥有彼此一致的温度的呢？

如果回想一下第4章，你或许会想到一种可能的解释：非定域量子纠缠除了可以使两个彼此间隔很远的粒子自旋关联之外，也可能会使间隔很远的空间区域温度关联。尽管这个想法听起来很有意思，但是别忘了第4章中的讨论是在完全受控的设定下进行，而空间强大的稀释能力会完全排除这种可能性。好吧，或许我们会找到更简单些的解释。或许很久以前，空间的各部分区域彼此非常接近，而这种彼此间的亲密接触使得彼此的温度也相同。这就好像把门开一会的话，原本温度很高的厨房和冰冷的卧室也会温度相同。但是，这种解释在标准的大爆炸理论中遭遇了失败的命运。我们现在来看一下为什么。

我们正在看这样一部电影，它讲述了宇宙从起源一直到今天的演化全过程。随便在什么时刻暂停一下电影，然后问问你自己：空间中的两块区域，就比如厨房和卧室吧，能够影响彼此的温度吗？它们能交换光和热吗？这种问题的答案取决于两个要素：区域之间的距离以及距离大爆炸的时间。如果两者之间的距离短于光在大爆炸之后的时间内可以行进的距离，那么它们就能够彼此影响，否则的话就不能。你现在可能会想，如果把我们的电影回放到足够接近大爆炸的时刻，可观测宇宙中的所有区域都会彼此影响；因为离大爆炸越近，不同空间区域的间隔也就越小，彼此也越容易相互作用。但是这样的推理太草率了，在这个思考过程中要考虑的并不仅仅是两者之间的距离是否足够的近，还要考虑留下来的时间是否足够的多。

我们现在来正确分析一下这个问题。慢慢回放这部关于宇宙的电影，我们要关注的是位于可观测宇宙两端的两块空间区域——这两块

区域之间的距离已经超出了它们各自的影响范围。我们要想把它们的间距缩小一半的话，就要把电影回放到少于一半的地方。但即使这样的话，两者之间的距离还是太远以至于不能彼此影响：它们之间的距离虽然减半，但是大爆炸到那个时刻的时间还不到大爆炸到今天的时间的一半，而光在这样的时间内能够行进的距离也没有光在大爆炸至今的时间内行进距离的一半多。类似地，我们可以一直这样通过回放电影来缩短空间中那两块区域的间距，但是我们也将发现，这会导致两者越来越难以彼此影响。在这样的宇宙演化中，两块区域在远古时期挨得越近，我们就越难——而不是越容易——想象它们会平衡彼此的温度。相对于光可以行进多远所带来的困难，回溯时间所导致的沟通困难显然更严重一些。

这就是标准大爆炸理论中的问题。在标准的大爆炸宇宙学中，万有引力只是一种吸引力；而且，自从宇宙开始以来，它就一直扮演着减慢宇宙膨胀的角色。对于减慢的事物来说，走过给定区域所花掉的时间要更多一些。比方说，秘书[49]迅猛地冲出门去，在2分钟内就跑完了前半圈，但今天它没在最佳状态，所以后半圈花了3分钟才完成。回放这场赛马的时候，我们必须往回放到一半多的时间（全长5分钟的影片我们要回放到开始2分钟的时候）才能看到秘书跑到半圈标记的位置。与此类似，在标准的大爆炸理论中，引力所起的作用是减慢空间的膨胀，因而，如果我们从任意时刻开始回放宇宙演化的电影，我们必须回放多于一半的时间才能使得空间中两块区域之间的距离缩减为一半。因而，如上所述，这就意味着虽然空间区域在更早的时刻变得更近，但彼此间也变得越难——而不是越易——相互影响，因而它们的温度相同这一事实也就令我们越困惑——而不是越明白。

物理学家们将某一区域所发出来的光在大爆炸之后的时间内所能到达的最远位置以内的空间区域定义为这一区域的宇宙视界（简称视界）。视界可以类比为我们在地球表面的某个制高点所能看见的最远距离，当然这只是类比。[15]视界疑难这个谜题基于这样一种认识：如

果空间中的两个区域的视界始终是分开的——永远不能彼此相互作用、交流，或者用任何方式彼此影响——那么它们怎么会有差不多完全一致的温度呢？

视界疑难并不是意味着标准的大爆炸模型不对，它只是需要解释。而暴涨模型就能提供这样的解释。

在暴涨宇宙学中有一个很短暂的时期，在这个时期内，排斥性的万有引力驱动着空间越来越快的膨胀。在关于暴涨宇宙学的电影中，你只需回放到少于一半的时间就能看到空间区域的间隔减半了。这就好比秘书花了2分钟的时间跑完前半圈后，状态正佳，又猛地提速，只用了1分钟的时间就冲完了最后半圈。于是在看回放的时候，你只需要把时间回放到比赛开始2分钟的时刻——你所回放的部分（1分钟）小于一半的时间（1分半）——就能看到秘书冲过半程标志的画面了。与此类似，空间中两块不同区域的间隔在暴涨时期急速增长，且增长的速度越来越快，这就导致在你看宇宙这部电影的时候，也只需要回放到少于——实际上少很多——一半的时间就能看到间隔减半了。这样一来，随着时间越往后，空间中不同区域也就越容易相互影响。因为随着时间越往后，不同区域彼此沟通的时间就会越多。计算表明，如果暴涨使空间膨胀的倍数达到了$10^{30}$以上的话，这样的数量级在大多数的暴涨模型中都可以实现，我们现在所能看到的所有空间区域——所有我们测过其温度的空间区域——都可以彼此沟通，就像相连的厨房和卧室之间的沟通那样简单，因而最初时刻的宇宙很快就会拥有相同的温度。[16]简而言之，暴涨阶段开始的时候，空间膨胀得很慢，整个空间的温度会变得一致，然后就进入了迅猛膨胀的阶段，一下子弥补了开始阶段的缓慢，临近的区域也彼此分开去了。

暴涨宇宙学就这样解释了弥漫于整个空间的微波背景辐射神秘的各向同性问题。

暴涨与平坦性疑难

暴涨宇宙学需要面对的第二个问题与空间的形状有关。在第8章中，我们强行使用了空间均匀对称的标准，找到了3种实现空间弯曲的方法。在两维的例子中，这3种可能性分别是正曲率（类似球面的形状）、负曲率（类似马鞍面的形状）以及零曲率（类似无限平坦的桌面或者有限尺度的电视游戏屏幕）。自从广义相对论建立以来，物理学家们就一直清楚这样一个事实：单位空间中的总物质或总能量——即物质或能量密度——决定了空间的曲率。如果物质或能量密度很高，空间就会把自己拉伸成球形，也就是说空间会有正的曲率。如果物质或能量密度很低，空间就会向外展成马鞍面，也就是说空间会有负的曲率。又或者像上一章讲过的那样，当空间中的物质或能量密度处于某一特殊值时——临界密度，相当于在每立方米的空间内有5个氢原子的质量（大约$10^{-23}$克）——空间将介于上述的两种极限情况之间，处于完美的平坦状态：也就是说，没有曲率。

现在我们来看看问题在哪。

掌控着标准大爆炸模型的广义相对论方程告诉我们，如果早期的物质或能量密度恰好精确的等于临界密度，它就不会随着空间的膨胀而发生变化，也就是说仍然保持在临界密度。[17]但是，即使物质或能量密度只比临界密度高一点点或者低一点点，接下来的膨胀过程都会使物质或能量密度远离临界密度。我们来感受一下具体的数字，如果在大爆炸后的1秒时，宇宙恰好比临界状态差了一点点，物质或能量密度达到临界密度的99.99%，那么计算就会告诉我们今天的宇宙中物质或能量密度将只有临界密度的0.00000000001%。这就像一个走在两面深渊的峭壁顶的登山者所面临的情况。每一步都落在峭壁顶的话就能安全通过。但是，稍稍迈错一步，不管是往右多一点还是往左多一点，倾斜都会瞬间放大，酿成不可挽回的后果（可能例子稍稍有点多了，不过还是很想讲一下。标准大爆炸模型的这个特点让我想起了很多年前大学宿舍的淋浴：如果你恰好调对了把手的位置，你就有舒服的温水洗澡。但你只要稍稍调错一点，不管是往哪边多偏了那么

一点，你就会被烫得发红或是被冻得发抖。所以那时候很多同学干脆就不洗澡了）。

物理学家们几十年来一直试图测准宇宙中的物质或能量密度。到了20世纪80年代，虽然离最终的结果还有一段距离，但是有一点已经是确定无疑的了：宇宙中的物质或能量密度既不比临界密度大成千上万倍，也不比临界密度小成千上万倍；这也就等于告诉我们，空间并没有充分的弯曲，既不是非常大的正曲率也不是非常大的负曲率。这样的观测事实难免令标准大爆炸模型有点难堪。人们认识到，标准大爆炸模型要想与实验观测相符，就必须找到某种机制——人们还不能找到这样的解释——从而使早期宇宙的物质或能量密度极端的接近临界密度。举个例子，计算表明，在大爆炸后的1秒，宇宙中的物质或能量密度同临界密度之间的差别不能超过临界密度的百万亿分之一；如果物质或能量密度和临界密度的差别稍稍大于这一极小的量，那么标准大爆炸模型所预言的今日物质或能量密度就会远远的大于实验观测值。因而，标准大爆炸模型中的宇宙就像在两面深渊的峭壁顶前行的登山者一样，走在极端窄小的山脊上。数十亿年前的微小差别将会导致今天的宇宙千差万别。这就是所谓的平坦性疑难。

我们已经讲过了平坦性疑难的关键之处，但还是很有必要再强调一下，这里的重要之处在于我们要理解平坦性疑难是一个问题。这并不等于说有了平坦性疑难就意味着标准大爆炸模型是错的。对于平坦性疑难，大爆炸理论的忠实信徒会耸耸肩无所谓地回答道：“宇宙当时的情况就是那样。”对于他们来说，只要好好地调节一下早期宇宙的物质或能量密度——从而使理论对今天的物质或能量密度预言与实验观测相符——平坦性疑难就不存在了。但是这样的回答让很多物理学家不舒服。如果一个理论，其成功要依赖于精细调节某些我们无法给予的基本解释的性质，那么这样的理论在物理学家看来就是极不自然的。由于标准大爆炸模型不能解释早期宇宙的物质或能量密度为什么非得精细调节到某个特定值，标准的大爆炸模型在很多物理学家看

来具有很强的人为性。因而，平坦性疑难可说是找出了标准大爆炸理论有极强的初始条件——我们不了解的远古时期的情况——敏感性这个弱点；平坦性疑难告诉我们，标准大爆炸模型只有假定宇宙必须如此，才能解释一些问题。

相反，物理学家们希望看到的理论是所能给出的预言不依赖于我们无法知道的量（诸如很久以前的情况如何）的理论。这样的理论可靠而自然，因为它们给出的预言并不依赖于各种很难甚至不可能直接确定的琐碎细节。暴涨理论就是一个这样的理论，它对平坦性疑难的解释很好地反映了这一点。

这里关键的一点，是要注意到吸引性的万有引力会放大任何背离临界物质或能量密度的偏差，而暴涨理论中的排斥性万有引力则正好相反：它可以减少对临界密度的偏差。要想直观地感受一下这一点，只需想想宇宙的物质或能量密度和其几何上的曲率之间的联系。特别需要注意的一点是，即使宇宙在早期具有高度弯曲的形状，暴涨膨胀之后，也会出现一片大到足以覆盖今天所看到的整个可观测宇宙的区域，而这样的区域看起来将是平坦的。事实上，我们早就熟知这种几何特性：篮球表面的弯曲可以一目了然；但是地球也是圆的就没有那么明显了，很多的思想家花了很多时间才理解了地球表面也是弯曲的这一事实。造成这种结果的原因在于，在一切不变的情况下，某一事物越大，它的弯曲就会变得越缓，其表面上的某一大小不变的区域就会看起来越平坦。如果你把内布拉斯加州[50]绕在直径几百千米的球面上，如图10.4（a）所示，内布拉斯加州看起来就是弯曲的。但在地球表面，所有的内布拉斯加人都会承认，该州看起来是平坦的。而你要是把内布拉斯加州放到比地球还要巨大的球体表面，它看起来就会更加平坦。在暴涨宇宙学中，空间被拉伸得如此之多，以至于我们所看到的可观测宇宙只不过是一个巨大宇宙的小小一部分。所以，如图10.4（d）所示的巨球上的内布拉斯加州一样，即使整个宇宙是弯曲的，我们可观测的宇宙也几乎是平坦的。[18]

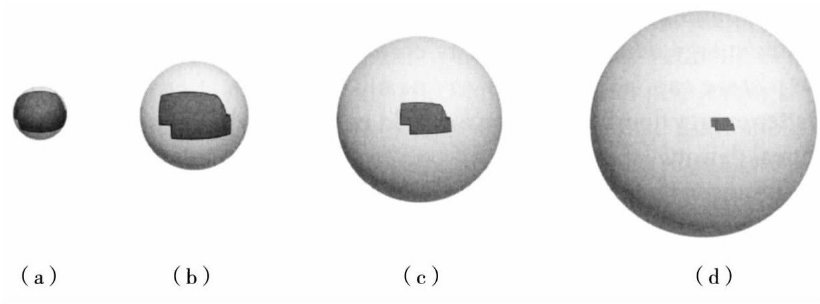



图10.4　图中大小固定的形状表示的是内布拉斯加州。可以看到，随着球的体积越来越大，其上的图形显得越来越平。这里，球代表的是整个宇宙，内布拉斯加州代表的是可观测宇宙——我们视界之内的宇宙。

这就好像在登山者的鞋底和峭壁顶上都安有强大的磁铁一般，一旦登山者朝着错误的方向迈上一步，磁铁之间强大的吸引力就会把登山者的脚拉回到山脊。与此类似，即使早期宇宙偏离临界密度一点，从而导致失去平坦性；暴涨膨胀也会将我们讨论的这部分空间拉回成平坦的形状，同时，我们所讨论的物质或能量密度也会被拉回到临界值。

进展与预言

暴涨理论对视界疑难和平坦性疑难的独到见解代表着巨大的理论进步。对于标准的大爆炸宇宙学来说，为了使演化到今天的宇宙既具有各向同性的特点，又具有同实验观测值相一致的物质或能量密度，它就必须对早期的初始条件进行准确的、未经解释的甚至有些怪诞的精细调节。人们当然可以假定真有这样的精细调节存在，就像标准大爆炸模型的追随者宣称的那样。但是这样一种毫无道理的办法使理论充满了人为性。而另一方面，暴涨宇宙学并不需要考虑早期宇宙物质或能量密度的种种细节。按照它自己的宇宙演化方式，暴涨宇宙学预言了我们观测到的宇宙应该近乎平坦。这也就是说，按照暴涨宇宙学的预言，我们观测到的物质或能量密度应该非常接近临界密度。

对于早期宇宙的细节不敏感可算是暴涨理论的一个优美之处，正

是因为有了这个性质，我们才能在对远古时期毫无了解的情况下给出确定的理论预言。但是我们也要问：这些预言是否能够经受得了精细的实验观测的检验呢？实验数据是否站在暴涨宇宙学一边呢？实验上是否观测到了暴涨宇宙学所预言的具有临界物质或能量密度的平坦宇宙呢？

多年的观测能告诉我们的答案只是“还不充分”。大量的天文学勘探仔仔细细地测量了宇宙中可见的物质或能量的总量，而这些勘探只得到了临界密度的5%。这样的观测结果显然不是标准大爆炸理论自然得出的极大密度或极小密度解释得了的——如果不考虑人为的精细调节的话——这也就是我在前面所说的，实验观测并没有发现今天宇宙的物质或能量密度比临界密度大成千上万倍或小成千上万倍。即便这样，5%离暴涨理论所预言的100%也还有很大的距离。但是，物理学家们早就认识到计算数据必须仔细，不能丢失任何可能性。天文学勘探发现的那5%只与能够发光的物质或能量有关，因为只有发光的能量或物质才能被天文望远镜看到。而早在几十年前，暴涨理论还没有诞生的年代，就已经有证据表明：宇宙很可能还有沉重的黑暗面。

黑暗预言

早在20世纪30年代，加州理工学院的天文学教授弗里茨·兹威基（一位著名的科学家，其刻薄的性格是出了名的。他非常喜爱对称性，甚至因此会将他的同事称为混球，他的解释是，无论从哪个角度看它们都是混蛋——所以是球形对称的混蛋，就是混球[19]）就认识到，后发座星系团（距离地球3.7亿光年的星系团，由上千个星系组成）的偏远星系中的可见物质移动得太快以至于无法聚集起足够的引力来使它们聚成团。兹威基通过分析指出，大多数快速移动的星系应该被甩出星系团，这就好像行驶中的自行车会甩出很多泥水一样。但我们却没有看到这样的现象。兹威基假定该星系团中还存在其他一些不能发光但却具有引力效应的物质，正是这些物质使该星系团聚在一起。通过计算，兹威基发现，要是这一解释正确的话，星系团质量的

绝大部分就应该由这种不发光的物质组成。1936年，威尔森山天文台的辛克莱尔·史密斯发现了确实的证据，通过对室女座星系团的研究，史密斯得到了类似的结论。但是，这两个人的工作，连同后继的一些工作，都存在着各种各样的不确定性，使得人们还没法确认就是大量的不可见物质的引力将星系聚成星系团。

接下来的30年间，有关不可发光的物质的实验观测证据越来越多，[20]但只有华盛顿卡耐基研究所的维拉·鲁宾，以及肯特·福特和其他少数人的工作才真正抓住了问题的关键。鲁宾和她的合作者研究了大量旋转星系中的恒星移动，他们的结论是：如果假定我们所看到的就是实际存在的全部，那么很多星系中的恒星将持续不断的向外飞去。这几位科学家的观测结果明确地告诉我们，可见的星系物质不可能产生出足够强的引力以控制住那些快速移动的恒星，使之不能自由飞散。详细的分析也同时指出，如果星系沉浸在不发光物质——其总质量需远远地超过星系中可见物质的质量——构成的巨大球状空间中（如图10.5所示），那么其中的行星就仍在引力的束缚之内。所以，就像有些观众只看到来回移动的白手套就能推断出魔术师隐藏在看不见的黑袍中一样，天文学家也能得出宇宙中充满着暗物质——那些不聚合成恒星因而不发光，所以除了借助于引力效应外没有任何办法知道它们存在的物质——的结论。而宇宙中的发光成分——恒星——仅仅是漂浮在暗物质之海中的灯塔。

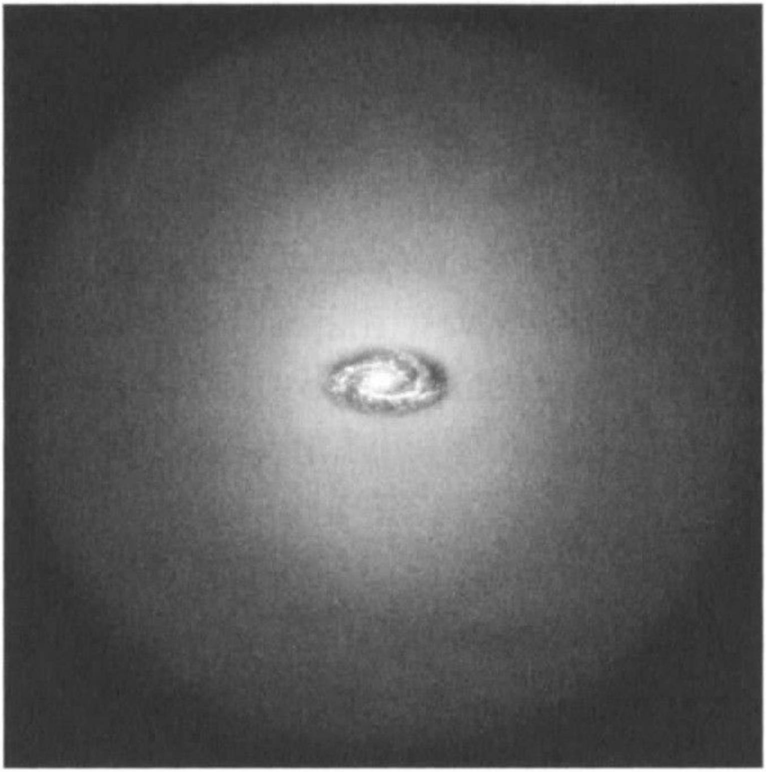


图10.5　浸在暗物质中的星系（为了使读者看清楚，图中高亮部分表示的是不可见的暗物质）。

如果必须有暗物质存在才能解释观测到的恒星和星系移动，那么，暗物质是由什么组成的呢？很遗憾，直到现在也没有人能给出肯定的答案。尽管天文学家和物理学家已经就暗物质是什么提出了很多种可能性——从各种奇异的粒子到小型黑洞，但暗物质究竟是什么仍然是天文学和理论物理学中的一个主要未解之谜。虽然不能确定暗物质究竟是什么，天文学家们还是能够通过准确的分析其引力效应来计算出暗物质在整个宇宙中究竟有多少。现在人们已经知道，暗物质大约占临界密度的25%。[21]再加上普通可见物质贡献的5%，我们已经找到了暴涨宇宙学所预言的物质或能量密度的30%。

好吧，虽然这已经算是很大的进展了，但科学家们还是得挠头，因为他们还是要知道那玩忽职守的70%哪去了（当然，这么说的前提是暴涨宇宙学真的正确）。但是，两组天文学家在1998年得到的结果出乎所有人预料，于是历史绕了一个圈，我们再次感受到了爱因斯坦是多么的富于预见性。

失控的宇宙

生病的时候，我们可能会听取不同医生的意见以确认自己的病情；物理学家们也是如此，如果他们得到的数据或理论带来了某些令人困惑的结果，物理学家们常常也会再从另外的角度看看同一个问题。如果从一个完全不同的观点出发也能得出与原始分析相同的结论的话，那么这种理论的可信度就很高。当从不同角度提出的解释最后都归结到一点的时候，我们就有理由相信这种解释已经切中要害。所以，很自然的，既然暴涨理论提出了某种异常怪异的东西——还没被观测到的那70%宇宙物质或能量密度——那么物理学家们必定希望依靠它来再确认一次暴涨理论。而很久以来人们就知道对减速因子的测量正好可以帮这个忙。

从暴涨开始的那一刻起，吸引性的万有引力就在减慢宇宙的膨胀，其减慢率称为减速因子。对减速因子的测量将是另一个能独立反映宇宙中物质总量的角度：更多的物质，无论发光与否，将导致更强的引力，因而会更明显的减慢空间的膨胀。

很多年来，天文学家们一直在试图测出宇宙的减速，这种测量虽然在理论上直截了当，但在实际中却是一个挑战。当我们观测远方的天体——比如星系或类星体——的时候，我们实际看到的是它们很久以前的样子：它们离我们越远，我们所看到的就是越早时期的天体。因而，如果我们能够测得它们以多快的速度远离我们，我们就相当于测到了在过去的某个时刻宇宙以多快的速度膨胀；然后，我们再测另外一个距离处的膨胀速度，这就相当于我们在测另外一个时刻的宇宙

膨胀速度。对比我们所测得的不同时期的宇宙膨胀速度，我们就会算出宇宙的膨胀在某段时间内减慢了多少，因而就求得了减慢因子。

所以，按这种方案测量减慢因子需要两个条件：一是要找到一种方法确定某一天体的距离（这样我们就可以知道这个天体在时间上距离我们有多远）；二是要找到一种方法测出该天体以多快的速度远离我们而去（这样我们就能知道在那个时候空间以多快的速率膨胀）。后一个要求很容易实现。远离我们而去的警车上的警笛音调变低；远离我们而去的天体上发出来的光的振荡频率也会变低。而且，诸如氢、氦、氧这样的原子——正是这些原子组成了恒星、类星体和星系——发出来的光已经在实验室中详细的研究过了，所以只要认真的对比我们从天体中测得的光的频率与实验室中的光的频率，我们就能得到某一天体离我们远去的速度。

但是前一个要求，找到一种确定某一天体距离我们多远的方法，则令天文学家们很是头疼。一般来说，我们会认为离我们越远的事物看起来越模糊，但是这一简单的事实却很难嫁接到定量测量上。你要想判断远方物体的亮度的话，就必须首先知道该物体的固有亮度——当它就在你附近的时候它有多亮。但是我们很难知道一个离我们几十亿光年远的物体的固有亮度到底是多少。一般的办法是找到一类天体，由于某种基本的天体物理原因，该类天体总是以准确固定的亮度发光。如果天空中缀满100瓦的灯泡，我们就好办了，因为这样的话，我们就可以根据灯泡的模糊程度来判断它到底离我们有多远（虽然看清很远地方的100瓦灯泡也是一种挑战）。但是宇宙空间并没有慷慨到这种程度；那么，我们该找什么东西来当标准亮度的灯泡呢？或者用天文学的话说，什么能起到标准烛光的作用呢？多年来，天文学家们已经研究了多种可能性，到目前为止最成功的一个选择是某一类特殊的超新星爆发。

当恒星用光它们的核燃料时，来自于恒星核心的核聚变的向外压力就会消失，恒星就会在其自身引力的作用下向内破裂。随着恒星核

心向自己坍塌，整个恒星的温度急速上升，有的时候就会导致一场规模巨大的爆炸，爆炸将会吹散恒星的外层，整个过程就像一场精彩的太空焰火表演。这种爆炸的恒星就是超新星；一颗超新星可以在几个星期的时间里始终保持着几十亿个太阳的亮度。它绝对是令人难以置信的天文奇观：单独一颗星星就会像整个星系那么亮！不同种类的恒星——不同大小，不同原子丰度，等等——会导致不同种类的超新星爆发。天文学家在很多年前就认识到某种超新星爆发总是具有相同的固有亮度，而这就是Ia型超新星爆发。

研究Ia型超新星首先要知道白矮星。白矮星——耗光了自己的核燃料但自身却没有足够的质量来触发一场超新星爆发的星体——会从附近的伴星中吸收表面物质。当白矮星的质量达到一个特定的临界值——大约是太阳质量的1.4倍，它就会通过核反应变为超新星。这样的超新星爆发总是发生在白矮星达到其临界质量的情况下，所以这种爆发的特性，比如其固有亮度等，总是大体上趋于同一个值。而且，因为超新星的亮度远远不是100瓦的灯泡所能比拟的，所以你不但可以有一个准确可靠的亮度源，还能在宇宙中够清楚地看到它。故而它们可以称为标准烛光的最佳候选者。[22]

20世纪90年代，两组天文学家分别由劳伦斯·伯克利国家实验室的索尔·派尔穆特和澳大利亚国家大学的布赖恩·施密特领导，决意通过测量Ia型超新星的远离速度测出宇宙的减速因子，从而测得宇宙的总物质或总能量密度。识别一次Ia型超新星爆发相对比较简单，因为超新星爆发所发出来的光总是急速增强然后又慢慢变弱。但是真正找到一颗Ia型超新星爆发却并不容易，因为在一个星系中，Ia型超新星爆发每几百年才出现一次。不过，通过利用视野广阔的望远镜同时观测数千个星系的这一革命性技术，两组天文学家在距离地球不同远近的地方发现了将近50次Ia型超新星爆发。耐心地从实验数据中计算出每一个超新星距离地球的远近及其远离速度后，两组天文学家得出了完全出乎意料的结论：宇宙在大约70亿岁后，膨胀率从未减速过。这

也就相当于说，膨胀率一直在加速。

这两个实验组发现，宇宙在其大爆炸后的70亿年间一直在减速膨胀，就像路遇高速收费站的汽车一样。人们所预期的正是这样的事实。但是，实验结果同时也告诉我们，就像一通过收费站的读卡器就踩油门的司机一样，宇宙的膨胀也在差不多70亿年的减速后开始加速。大爆炸之后的第70亿年的膨胀率小于第80亿年的膨胀率，而第80亿年的膨胀率又小于第90亿年的膨胀率，依此类推，以前任意一个时期的膨胀率都要小于今天的膨胀率。人们所预期的空间减速膨胀变成了完全出乎意料的加速膨胀。

这究竟是怎么一回事呢？这个问题的答案就与物理学家们曾经问过的那丢失的70%的物质或能量密度有关。

丢失的**70%**

现在想想1917年和爱因斯坦引入的宇宙常数，你所拥有的信息足以让你明白宇宙加速是怎么回事。普通的物质和能量产生普通的吸引性万有引力，这种引力会减慢空间的膨胀。但是随着宇宙的进一步膨胀，宇宙间的事物分散得更加厉害，因此使膨胀减速的引力就变得更弱了。这一下事情出现了转机。如果真的有一个宇宙常数的话——并且它的值又小得恰到好处的话——那它所产生的排斥性万有引力在大爆炸之后的70亿年内就会被普通物质所产生的吸引性万有引力盖过，其净效果就是减慢宇宙的膨胀，而这一点正与实验数据相符。随着时间的流逝，普通物质分散得更开，其所产生的引力也就更弱；这时，宇宙常数（其值并不因为物质分散而有所变化）所产生的排斥性万有引力就会逐渐占据上风，于是减速膨胀的时期让位于加速膨胀的时期。

于是，在20世纪90年代后期，根据上述的推理和对数据的深入分析，派尔穆特组和施密特组都认识到，爱因斯坦在80年前为广义相对

论方程引入宇宙常数并不是一个错误。他们认为宇宙的确应该有一个宇宙常数。[23]但是他们并不认同爱因斯坦提出的宇宙常数大小，其原因在于爱因斯坦所追求的是排斥性万有引力和吸引性万有引力彼此相消从而导致静止的宇宙，而派尔穆特和施密特等人需要的是在某些年间排斥性的万有引力要居于主导地位。虽然派尔穆特和施密特等人的发现还需要进一步的实验确认以及更加仔细的分析，但我们必须得说，爱因斯坦再一次预见了宇宙的基本性质，只不过这次足足花了80年才得以实验验证。

超新星的远离速度取决于普通物质的吸引性万有引力和由宇宙常数所提供的“暗能量”的排斥性万有引力之间的差别。如果将普通物质的总量——可见不可见的都算上——算作临界密度的30%的话，超新星研究者们发现，宇宙常数的暗能量必须贡献出70%的临界密度才能使超新星的远离速度满足实验观测。

这是一个令人印象深刻的数字。如果真的是这样，那么我们就不仅仅有普通物质带给宇宙的那无足轻重的5%的物质或能量，以及现在还无法确知其成分的暗物质带给宇宙的25%的物质或能量，我们还将拥有遍布于整个空间的宇宙常数所导致的那全然不同又神秘莫测的暗能量所贡献的宇宙的大部分物质或能量。如果这些想法真的正确的话，那么哥白尼的学说将会神奇的扩充：不但我们所在的位置不是宇宙的中心，就连组成我们的物质也只是沧海一粟。即使将质子、中子、电子全都从宇宙中拿走，宇宙的总质量或总能量也不会减少多少。

除此之外，70%这个数字还有另外一个令人印象深刻的原因。将宇宙常数贡献的70%和普通物质及暗物质贡献的30%合在一起的话，宇宙的质量或能量正好100%的符合暴涨理论的预言！因此，超新星的数据所展现出来的外推力正好可以用暗能量来解释；而要正确的解释超新星数据，暗能量就要正好贡献出宇宙中看不见的那70%的质量或能量；而这丢失的70%正好是一直令暴涨宇宙学家挠头的问题。所

以，有关超新星的实验测量与暴涨宇宙学形成了美妙的互补关系。两者彼此印证，分别是对方的佐证。[24]

将观测结果与暴涨理论组合起来，我们就得到了如图10.6所示的宇宙演化图像。极早期的时候，宇宙的能量存在于暴涨子场中，这时的暴涨子场并没处于最低能量态。由于负压的作用，暴涨子场驱动了一场暴涨膨胀。在大约$10^{-35}$秒之后，暴涨子场滑落到其势能碗的最低处，暴涨阶段结束，暴涨子场释放出了蕴藏的能量，这些能量转化成了普通的物质和辐射。之后的几十亿年间，这些常见的物质所释放出来的吸引性万有引力使空间的膨胀不断减速。但是随着宇宙逐渐变得稀薄，这种排斥性的万有引力慢慢减小。时间到了距今大约70亿年前，普通的吸引性万有引力弱到了一定程度，这个时候来自于宇宙常数的排斥性万有引力开始居于主导地位，于是空间的膨胀率开始上升。

距今大约100亿年的时候，除了离我们最近的一些星系，大部分星系都随着空间的膨胀而离我们远去，它们的速度快到超越了光速，所以我们无论使用什么样的望远镜都无法看到那些飞走的星系。如果这些理论真的是正确的话，那么到了遥远未来的某一天，宇宙将会变成巨大、空荡、毫无生气之所。

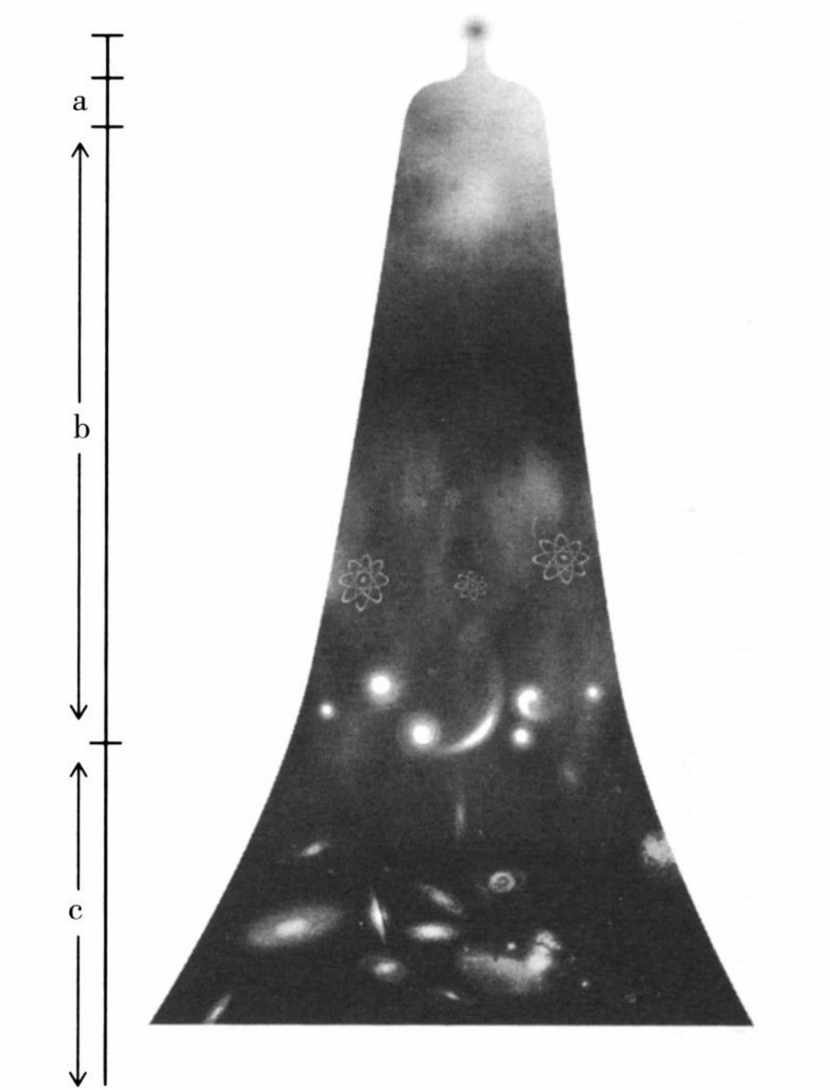



图10.6　宇宙演化的时间线。a：暴涨。b：标准的大爆炸演化。C：加速膨胀时期。

谜题与进展

有了这些发现之后，看起来我们已经解决了宇宙学中的种种疑难。标准大爆炸宇宙学不能回答的问题——是什么使空间向外膨胀？

微波背景辐射的温度为什么一样？为什么空间看起来是平坦的？——被暴涨宇宙学一一解答。但即使这样，有关基本起源的困难问题仍然存在：暴涨之前是否还存在着某个时期，如果是的话，这一时期的宇宙是什么样子？是什么使暴涨子场离开它的最低能量态从而触发了一场暴涨？还有一个最新的问题，宇宙为什么会由这些好不般配的成分——5%的普通物质，25%的暗物质，70%的暗能量——组成呢？要是不考虑这样的宇宙成分极好的符合了暴涨理论关于宇宙要有100%的临界密度的预言，也不考虑暴涨理论还同时解释了在超新星的研究中所发现的宇宙膨胀的话，大多数物理学家都会觉得这种大杂烩似的组合毫无吸引力可言。很多人都会问：宇宙为什么会有这么复杂的组成呢？为什么这些全然不同的成分的比例会如此随机呢？是不是有什么深层理论可以解释这些问题呢？

现在还没有人就这些问题提出令人信服的解释。正是这些问题和另一些紧迫的研究问题推动着当前的宇宙学研究；这些问题的存在时刻提醒着我们，在宣称我们已经完全理解了宇宙的起源之前必须打开这些混乱的结。不过，要是不考虑遗留的这些艰巨挑战，暴涨理论可算是当前最为成功的宇宙学理论。固然，物理学家们对暴涨理论的信心基于我们在本章中讨论过的那些成功之处；但是，暴涨理论并不仅仅有这些成功之处，人们对它的信心随着越来越多的发现而越来越足。我们将在下一章中看到，许多其他考虑因素——既有来自实验观测上的也有来自理论上的因素——同样会使工作在这一领域的物理学家们相信，暴涨理论是我们这一代物理学家对宇宙学所做的最重要、最持久的贡献。

## 第11章　缀满钻石的天空中的量子[51]

暴涨、量子涨落与时间之箭

暴涨理论的发现开启了宇宙学研究的新纪元，在那以后的20年间，关于这一方面的研究论文数以千计。科学家们搜遍了这个理论的

犄角旮旯，他们的细致程度远超你的想象。尽管很多科学家在这方面的工作更多关注的是技术上很重要的细节；但还是有一些科学家走得很远，他们发现，暴涨理论并不仅仅能用来解释标准大爆炸宇宙学力所不及的那些宇宙学问题，还能够为很多老问题带来强有力的新办法。在这些老问题中的3个——星系之类的团状结构的形成，造就我们今日所见的宇宙究竟需要多少能量，以及（在我们的故事中最重要的一个问题）时间之箭的起源——上暴涨理论取得了实实在在的，按某些人的说法甚至是令人叹为观止的进展。

让我们一起来看看。

用量子语言写的空中文字

暴涨宇宙学对视界疑难和平坦性疑难的解释只是其盛名的开始。我们已经看到，这两个成就曾是它的主要贡献。但是随着时间的推移，越来越多的物理学家相信暴涨理论的另一项成就的重要性可能不亚于其在视界疑难和平坦性疑难这两个问题上的贡献。

这个重要的成就与一个到目前为止我还没有提请你注意过的问题有关，这个问题就是：星系、恒星、行星，或者其他一些宇宙中的团状结构究竟是怎么形成的呢？在前面的3章中，我们关注的一直是宇宙学意义上的大尺度，在这样的尺度上宇宙表现出来的是同一性；这一尺度如此之大，以至于我们可以将一个星系看做是一个$H_2O$分子，而将整个宇宙看做一杯水。但是，宇宙学的研究不可能永远不触及星系之类的团状结构，随着我们的研究尺度变得越来越“精细”，我们早晚会问到有关星系的问题。于是，我们又遇到了一个谜题。

如果宇宙真的光滑均匀且在大尺度上各向同性的话——这是宇宙学的观测告诉我们的事实，所有有关宇宙学的理论其核心必是如此——那么小尺度上的团状结构是怎么出现的呢？标准大爆炸宇宙学的忠实信徒对这个问题不屑一顾，他们再一次用精细调节宇宙的初始条

件来解决问题，他们会这样说："极早期的宇宙大体上是均匀各向同性的，但是这种均匀性并不完美，在某些地方会有小小的疙瘩存在。至于为什么会这样，没有人知道，但当时的确就是如此。随着时间的推移，这些小疙瘩会逐渐变大，因为它们会比附近的环境具有更强的引力，因此会慢慢地把物质聚拢过来而变得越来越大。最终，这些小小的疙瘩变成了恒星、星系这样的结构。"乍看之下，这样的解释还算合理，可惜它有两大缺陷：这种说法既没能够解释初始时候为什么会有大体上的均匀性，也没能够解释为什么还有小小的但却非常重要的不均匀性。这两个问题的存在，正好给了暴涨理论大显身手的机会。在前面的章节中我们已经讨论过暴涨理论是如何解释大尺度上的均匀性，现在我们来看看暴涨理论是如何对付另外一个问题的。在暴涨宇宙学中，最后导致恒星与星系形成的初始不均匀性来自于量子力学。

这一重要的思想来自于两种貌似毫不相干的物理学理论——空间的暴涨膨胀和量子力学的不确定原理——的相互影响。不确定原理告诉我们，宇宙中各种互补的物理性质的不确定度之间总有某种微妙的平衡。我们最熟悉的例子与物质有关（参见第4章）：我们将一个粒子的位置测得越精确，对该粒子速度的测量就会越不精确。除物质外，不确定原理还可以应用于场。类似于对物质的有关讨论，我们也可以根据不确定原理知道，空间中某一位置处某种场的值测得越准，该场在这点的值的变化率就测得越不准（在量子力学中，粒子的位置与位置的变化率——速度——构成一对互补关系，同样的，空间中某点的场值与该点处场值的变化率也构成一对类似的互补关系）。

我喜欢这样总结不确定原理：简单地说，量子力学使事物躁动不安。如果我们不能完全搞清楚一个粒子的速度，我们就没法知道它在下一时刻的位置，因为这一时刻的速度决定下一时刻的位置。在某种意义上说，粒子的速度可以取任意值；或者更确切地说，粒子处于多种速度的混合状态，因而粒子会处于毫无规律的狂乱运动状态。场的

情况与此类似，如果我们不能将场值的变化率完全确定下来，我们就确定不了场值在下个时刻的大小。在某种意义上说，场以或这或那的速度上下起伏；或者更准确地说，我们可以假定场处于各种各样的改变率的混合状态，因而其值处于混乱随机的涨落之中。

在日常生活中，我们完全感觉不到粒子或场的量子涨落的存在；之所以如此是因为量子过程发生在亚原子尺度上。而我们知道，暴涨对亚原子尺度有很大的影响。突然而至的暴涨膨胀将空间拉伸到一个大得难以想象的程度，原本存在于微观尺度上的一切突然暴露在宏观尺度上。暴涨宇宙学的先锋们[1]率先认识到，不同空间位置的量子涨落之间的随机差别会导致微观尺度上存在极小的不均匀性；由于无迹可循的量子效应，某一位置的总能量可能会和另一位置的总能量有所不同。于是，到了空间暴涨的时候，这些微小的差别就会被放大到量子尺度之外，从而导致团状结构出现；这就好像我们吹大了气球的时候会看到上面本来看不清的小图案。物理学家们相信，标准大爆炸理论的忠实信徒无法证明而只是简单的归结为“当时就是那个样”的团状结构起源正是上面解释的那样。暴涨宇宙学用不可避免的量子涨落的放大效应解释了一切：暴涨将量子涨落所带来微小不均匀性放大到整个星空。

短暂的暴涨时期过后，这些小小的团状结构在接下来的几十亿年间由于引力的作用而继续增长。就像标准的大爆炸宇宙学描述的那样，由于团状结构比其周围的环境更加密实，因而有更强的引力，它们可以将周围的物质慢慢吸引过来，从而使自身变得更大。最后终于有一天，这些团状物质大到了一定程度，恒星、星系也就形成了。毫无疑问，小小的团状结构最终发展成星系要经历数不清的详细步骤，而我们对于其中的很多仍不了解。但是我们已经了解了整体框架：量子世界中，由不确定原理带来的量子涨落使得世间万物都不能保持完美的均匀性。对于经历过暴涨的量子世界来说，这种小尺度上的不均匀性将被放大成大尺度上的不均匀性，这种不均匀性会在未来的某一

天导致星系之类的天体的形成。

基本思想就是这样，不想深究这个问题的读者可以略过下面的几段。但对于那些感兴趣的读者，我很愿意再进一步深入探讨一下。我们首先回忆一下前面讲过的内容：暴涨阶段结束的时候，暴涨子场的值将落到其势能碗的最低位置，蕴藏在暴涨子场中的能量和压强将全部释放出来。我们曾经说过，在整个空间中情况都是如此——暴涨子场不管在这里还是那里经历的都是相同的演化过程——这是我们从方程中自然得出的结论。不过，这种情况仅当我们略去量子效应时才严格成立。平均来说，暴涨子场的值的确会跌落到势能碗的最低位置，就像球会从斜面上滚落一样。但是，就像从碗中滑下来的青蛙可能会到处乱蹦，暴涨子场也会由于量子力学的不确定原理而处于随机涨落的状态。在场值降低的这个过程中，某些地方的值可能突然升高一下，另一些地方的值又可能突然比周围的值降得更低一点。正因为有这种量子涨落存在，不同位置的暴涨子场可能会在不同的时刻降至最低能量。也就是说，暴涨过程在空间中不同位置的结束时间可能有所不同，这样的话，不同位置处的空间膨胀率就有可能有微小的差别，由此造成的不均匀性——就像褶皱一样——有点类似于比萨师傅揉面团时着力不均而造成的凸包。人们一般认为，相比于天文学尺度，来自于量子力学的涨落通常会因为太小而微不足道。但是在暴涨理论中，空间的膨胀倍数如此之大——每$10^{-37}$秒就能膨胀1倍，以至于临近位置在暴涨时间上的微小差别都会导致巨大的空间褶皱出现。事实上，人们在某些具体的暴涨模型中所做的计算表明，按这种机制产生的不均匀性可能有点太大了；所以有的时候研究人员非得仔细调节相关参数不可，要不然的话，理论所预言的宇宙可能会显得比实际情况臃肿得多。总之，暴涨宇宙学的确为我们带来了一种现成的机制，使我们得以理解小尺度上的不均匀性如何使得一个在最大尺度上看似均匀的宇宙中出现了恒星和星系这样的团状结构。

暴涨宇宙学告诉我们，遍布于整个天空如钻石般闪亮的那一千多

亿个星系不是别的，正是量子力学在天空中的自我展示。对于我来说，这样的认识简直就是现代科学中最伟大的奇迹。

宇宙学的黄金时代

用卫星对微波背景辐射温度所做的细致入微的观测为上节讨论的那些思想提供了奇迹般的证据。我已经反复强调过，天空中各处的辐射温度彼此符合得很好。但我没有提过的是，这种符合只能到小数点后的第4位，实际上，不同位置的辐射温度有微小的差别。精确的实验观测——最早由1992年的COBE（宇宙背景探测器）完成，近年来的WMAP（威尔金森微波各向异性探测器）也贡献不菲——告诉我们，空间中某处的温度可能是2.7249开，另外的地方可能是2.7250开，还有的地方可能是2.7251开。

奇妙的是，天空中这种极其微小的温度变化竟有规律可循，而这种规律性又可以用解释天体形成的同一机制——暴涨过程将量子涨落放大——加以解释。简单地说就是，微小的量子涨落遍穿天际，这种涨落使得空间中的某处可能稍热一点，另一处又稍冷一点（来自于稍密实些区域的光子必须用更多的能量来克服引力的作用；因此，相对于来自稍疏散区域的光子，这些光子的能量和温度要稍微低点）。以这样的观点为基础，物理学家们做了精细的计算，并根据计算结果预言了微波辐射温度随位置变化而变化的情况，如图11.1（a）所示（细节并不重要。图中横轴所示的是天空中两点的角度差，水平轴表示的是两点间的温度差）。图11.1（b）表示的是实验数据与理论预言之间的对比，其中黑点表示的是卫星上得来的观测数据，我们可以看到，理论与实验符合到令人难以置信的程度。

我希望你已经被实验数据和理论预言的符合程度深深触动，因为如果不是那样的话，就意味着我没能很好地传递出这一结果的神奇之处。所以，为以防万一，我还是再来强调一下这到底意味着什么。近年来，人们利用安装在卫星上的望远镜测量了微波光子的温度，这些

光子已经朝我们飞了大约140亿年，期间并未受到任何阻碍。人们发现，来自于空间中不同方向的光子具有几乎一样的温度，这些光子在温度上的差别只有千分之几的量级。而且，人们从观测数据中发现，来自于不同方向的光子之间的微小差别具有某种特定的规律，这种差别上的规律性可以用图11.1（b）中的黑点表示。而最神奇的地方在于，今天的人们可以在暴涨理论的框架下利用理论计算来解释这些微小的温度差——要知道，这种温度差可是在140亿年前就确定下来的——的规律性；而且，追本溯源，我们发现，这种解释的关键竟与量子力学的不确定原理有关。太神奇了，难道不是吗？

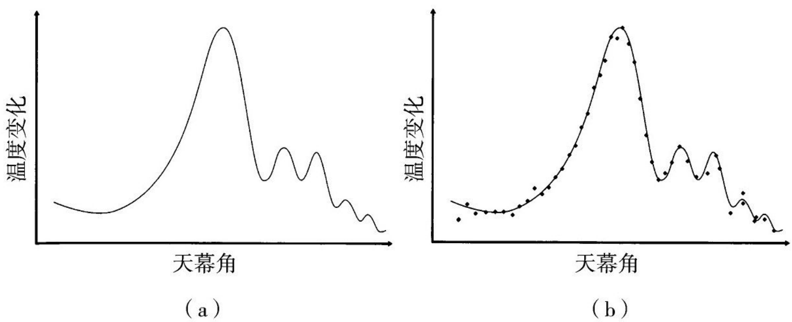



图11.1　（a）：暴涨宇宙学所预言的微波背景辐射在天空中不同位置处的温度变化。（b）：用来与理论预言对比的卫星实验观测数据。

这一成功使很多物理学家相信暴涨理论的有效性。而且，同样重要的是，这样或那样的精确天文学观测（近年来才逐渐得以实现）使宇宙学从只能依靠假想与猜测的时期逐步过渡到有坚实实验基础的成熟时期，一个令工作在这一领域的科学家们兴奋万分的黄金时代即将到来。

## 创造一个宇宙

有了这样的一些进展，物理学家们已经等不及要看看暴涨宇宙学到底能走多远了。比如说，物理学家们很想知道，暴涨理论是不是能

够回答莱布尼茨所提出来的终极问题——为什么会有宇宙存在？以我们现在的理解水平，这个问题可能太大了。即使某一宇宙学理论真的在这个问题上有所进展，我们也可以问为什么会有这样一种特别的理论；它的假设、参数、方程都是哪来的？这样的话就仅仅算是把这个终极问题又往后推了一步。除非只依靠逻辑本身我们就能够要求宇宙存在，并且要求宇宙只能被唯一的一组方程和参数掌控，这样的话或许我们才会满意。但到目前为止，这还只能算是一场白日梦。

与此相关但却没这么含糊不清的一个问题在各个年代都会被人们问到，这个问题就是：组成宇宙的质量或能量究竟来自何方？对于这个问题，暴涨理论尽管不能完全回答出来，但却能带来一些新的灵感。

在探讨这个问题之前，我们先来想象一个巨大但却柔软的盒子，这个盒子中满是跑跳不休的小孩子。这个盒子密封得很好，绝对不会散发热量或能量。但是整个盒子很柔软，它的墙壁甚至可以向外移动。孩子们不停地撞向墙壁——每次几百个孩子一起撞过去，一拨之后马上就是另一拨——这样一来盒子就会慢慢膨胀。于是，你可能会想，由于盒子完全密封，孩子们的总能量会完全驻留在盒子内。要不然的话，那些能量还能去哪呢？这样的想法虽然貌似合理，但实则不然。能量还是有地方去的。孩子们在每次撞击墙的过程中都会消耗能量，这些能量的大部分都转化成了墙的移动。正是盒子的膨胀吸收消耗了孩子们的能量。

现在，孩子中的调皮鬼们打算换个玩法。他们用大量的橡胶带将彼此相对的两面墙钩住。这样一来，橡胶带就会对墙产生一个向内的负压力，这种负压力的作用效果正好与孩子们撞击墙的作用效果相反；这样设置的橡胶带不但没有将能量传递给盒子用以膨胀，相反的，橡胶带的负压还能“消化”膨胀的能量。随着盒子的膨胀，橡胶带也越拉越紧，而橡胶带越紧就意味着蕴藏于其中的能量越多。

当然，我们真正感兴趣的并不是膨胀的盒子，而是膨胀的宇宙。充满空间的也并不是孩子们和数不清的橡胶带；根据我们的理论，在宇宙演化的不同时期，空间中充满的是均匀的暴涨子场或是普通物质粒子（电子、光子、质子，等等）。但是，简单分析一下之后，我们就可以将从盒子的故事中得出的结论应用于宇宙。快速运动的孩子们会承受盒壁由于膨胀而向内施加的力，宇宙中快速运动的粒子也会承受宇宙由于膨胀而施加的向内的力：这种向内的力就是万有引力。根据这段分析（数学上也是如此），只要将万有引力替换为盒子的墙，我们就可以将宇宙类比为盒子。

因而，正如孩子们所拥有的总能量会在盒子膨胀的时候不停地转变为墙的能量而损失；普通的物质粒子和辐射所携带的能量也会在宇宙膨胀的时候不断地转化为引力而损失。此外，正如调皮鬼们的橡胶带会在膨胀的盒子内施加负压；均匀的暴涨子场也会在膨胀的宇宙内施加负压。所以，就像橡胶带的总能量会随着盒子的膨胀而增加，因为橡胶带可以从盒子的墙中抽取能量；暴涨子场的总能量也会随着宇宙的膨胀而增加，因为它可以从引力中获得能量。[52]

总而言之：随着宇宙膨胀，物质和辐射的能量遗失给引力，而暴涨子场则从引力那里获得能量。[53]

当我们试图解释构成星系、恒星或者宇宙间其他天体的物质和辐射的起源时，这种看法的重要性就会变得非常明显。按照标准大爆炸理论的说法，物质和辐射所具有的质量或能量会随着宇宙的膨胀而减少，因此，早期宇宙中的物质或能量远比今天我们所能看到的多。但是这样一来的话，标准大爆炸理论不但没能解释宇宙中现存的所有物质或能量的起源，还使自身陷入到一场永无希望取胜的战斗中：越是早期的宇宙中，存在着越多的质量或能量等着标准大爆炸理论解释。

但是在暴涨理论中，一切恰恰相反。还记得吗？根据暴涨理论，物质和辐射产生于暴涨阶段的末期，这个时候的暴涨子场从势能碗的

高处跌落到最低能量处，并且释放出所蕴藏的全部能量。因而，人们就要问暴涨理论是否能够解释这样的问题：在暴涨阶段结束的时候，暴涨子场如何才能产生足够多的物质或能量以符合今日宇宙中的物质和辐射的量。

这个问题的答案是：暴涨理论不费吹灰之力就能做到这一点。我们之前已经解释过，暴涨子场就像引力的寄生虫一样——以引力为食——所以暴涨子场所具有的总能量会随着空间的膨胀而增加。或者更准确地说，数学分析表明，暴涨子场的能量密度在整个暴涨膨胀时期保持不变，这就意味着暴涨子场所具有的总能量正比于暴涨子场弥漫于其中的空间的大小。我们在前面的章节中已经看到，宇宙的尺寸在暴涨时期将至少膨胀$10^{30}$倍，这意味着宇宙的体积将至少膨胀（$10^{30}$）$3=10^{90}$倍。所以，暴涨子场所具有的总能量增加了同样的倍数——$10^{90}$倍：当暴涨阶段趋于终结的时候，也就是在暴涨开始后的大约$10^{-35}$秒，暴涨子场总能量至少增加了$10^{90}$倍。这就意味着，一旦暴涨过程启动，即使暴涨子场最初所拥有的能量并不很多，它也会由于膨胀而将自身所拥有的能量放大到极大的地步。简单的计算告诉我们，即使暴涨子场最初填满的空间长度只有$10^{-26}$厘米，其重量不过20磅，它通过暴涨膨胀获得的能量也将多到足以解释今日宇宙间的一切物质之起源。[2]

因而，不同于标准大爆炸理论所给出的解释——这一解释将导致早期宇宙中的总物质或总能量多到无以言表；暴涨宇宙学，从最初微小的空间上以20磅的暴涨子场起家，通过不断的“挖掘”引力，制造出了今日宇宙的全部普通物质和辐射。当然，这绝不意味着暴涨理论回答了莱布尼茨的终极问题——为什么是存在一个宇宙而不是相反，因为我们还没能够解释为什么会有暴涨，甚至为什么会有空间存在。但是我们无法利用暴涨理论解释的东西已经只有20磅——甚至比我的狗洛基还轻，而这一点显然是标准大爆炸理论根本做不到的。[54]

## 暴涨，平滑性与时间之箭

或许我的狂热早已出卖了我，或许已经有读者们感觉到我的偏心了；但是说真的，在所有的当代科学进展中，人类在宇宙学领域所取得的成就最令我敬畏。看起来我永远都无法忘怀许多年前第一次读广义相对论时的那种迫不及待的心情了；就是在那个时候，我第一次认识到，原来我们可以利用广义相对论，从时空中我们所在的这小小角落出发，来了解整个宇宙的演化。几十年后的今天，技术的进步已经将过去只能抽象讨论的早期宇宙演化问题付诸实践检验，而我们也已看到，理论竟然真的有用。

还记得吗？我们曾在第6章和第7章中讨论过，宇宙学的研究除了能令我们知晓时间和空间故事外，还能给我们以另外的启迪：帮助找到时间之箭的源头。通过那几章的讨论我们发现，关于时间之箭唯一可靠的一点是早期宇宙必定高度有序，也就是说，早期宇宙的熵必须极低；这个时期的熵将决定未来的熵的量，而宇宙必定会向着熵增的方向演化。要是人们根本就没有按照顺序装订好《战争与和平》这本书，那么书的页码就没法被弄乱，因为它本来就是乱的；同样的，早期宇宙若不是处于高度有序的状态，后来的宇宙所处的状态也没法变得更加无序——比如牛奶溅落、鸡蛋破碎、人们变老这样的无序状态——因为它本来就处于无序状态。于是我们的问题来了，高度有序的低熵状态究竟是怎么开始的呢？

在这个问题上，暴涨宇宙学可以派上用场。但在我们深入讨论之前，先让我来更加详细地说一说问题本身，以防止某些重要的细节之处逃过了你的眼睛。

因为有过硬的证据存在，大部分人都相信这样一点：在宇宙的早期历史中，物质均匀地遍布于空间。但是一般来说，这就意味着早期宇宙处于高熵状态——就像可乐瓶中的二氧化碳分子均匀地分散于房间内一样——而我们则很难解释这一点。不过，如果我们将引力也考虑进来的话——讨论整个宇宙的时候当然应该考虑引力——我们就得说，物质均匀分布的状态实际上是一种高度有序的低熵状态，因为引

力原本会使物质聚团。与此类似，光滑均匀的弯曲空间也处于一种低熵的状态；相比于崎岖不平的弯曲空间，光滑的弯曲空间可算是高度有序（这就好比使《战争与和平》的页码有序的编号只有一种，但是使其页码混乱的编号却有很多种；空间也是这样，崎岖不平怪模怪样的空间形状可有很多种，有序、光滑、均匀的空间形状可并不多）。于是就有一道难题留给了我们：为什么早期宇宙会处于一种物质均匀分布的低熵（高度有序）状态，而不是物质聚团分布——就像各种各样的黑洞——的高熵（高度无序）状态呢？为什么空间会异常准确地按照光滑、有序、均匀的形状弯曲，而不是按照皱皱巴巴——也好比黑洞那样——的形状弯曲呢？

保罗·戴维斯和堂·佩奇首先就这个问题进行了仔细的研究。[3]他们发现，暴涨宇宙学在这一研究中发挥了至关重要的作用。要看清这一事实，我们首先得记住这样一点，我们所讨论的这一问题基于这样一个假设，即物质一旦在某处聚团，就会形成较大的万有引力，从而吸引更多的物质聚拢过来；对应说来，空间中的某处一旦形成褶皱，就会产生更大的引力，从而使得褶皱更为严重，进而导致空间高度不均匀的弯曲。总之，将引力也考虑进来的话，物质聚团、空间出现褶皱才是高熵的状态。

但是我们也得知道，上述的推理过程依赖于这样一个假设：普通的万有引力是一种吸引力。物质聚团也好，空间出现褶皱也罢，它们的出现都是因为有更强的万有引力，能够把周围的物质吸引过来。但是暴涨阶段的万有引力是一种排斥性的力，所以这一阶段的情况会有所不同。比如说空间形状。排斥性万有引力那巨大的外推力会驱使空间迅速膨胀，从而拉平初始时的褶皱；这一过程有点像给气球吹气：开始时皱皱巴巴的气球，吹足了气后就会鼓起来，从而变得光滑了。
55而且，空间的体积在暴涨时期膨胀的倍数非常之巨大，因而使得聚团物质的密度大大减小了；我们可以这么想象这一情况：或许你养的鱼数目比较多，而你的鱼缸又不大，因而看起来那些可爱的小东西总

是挤来挤去的，但如果把你的鱼缸换成奥运比赛所用的游泳池，那么这些小鱼就再也不会感到拥挤了。总之，虽然吸引性的万有引力使得物质聚团，空间出现褶皱；但是排斥性的万有引力则正好相反：它会消除这些效应，使宇宙变得光滑、均匀。

因而，暴涨时期结束的时候，宇宙的尺度变得难以想象的巨大，空间中的褶皱都被拉平了，原本聚团的物质也被冲散了。而且，随着暴涨子场滑落到势能碗的最低处，暴涨膨胀接近尾声，暴涨子场所蕴藏的能量将会释放出来，在整个空间中均匀地填满普通的物质粒子（虽然直到很小的尺度这种均匀性仍然存在，但是在更小的尺度上会出现量子涨落带来的不均匀性）。总之，我们已经前进了一大步。我们通过暴涨理论所得到的结果——均匀地分布着物质的光滑而均匀的空间——正是我们试图解释的东西。要想解释时间之箭，我们需要的正是这种低熵的初始状态。

熵与暴涨

事实上，这一进展意义非凡。但是有两个问题还没有弄清。

首先，看似我们得到了这样的结论：暴涨膨胀使空间变得光滑，物质分布变得均匀，所以暴涨过程中总的熵似乎降低了，但这就意味着出现了某种破坏热力学第二定律的物理机制——而不是统计上的意外那么简单。要真是这样的话，那么不是热力学第二定律错了就是我们的推理错了。但实际上，我们并不需要面对这样的两难选择，因为总的熵并没有随着暴涨过程的发生而有所下降。真实的情况是，暴涨发生的时候，总的熵会有所增加，只不过增加的总量比我们原本以为的量要少得多。你已经知道，暴涨阶段结束的时候，空间会被拉扯得非常光滑，因而引力对熵的贡献——物质越是聚团、无序，空间越是不光滑，这种熵就越大——达到最小值。但是，暴涨子场在滑到其势能碗最低位置的时候，会释放出巨大的能量，这些能量会产生出数目惊人的物质——大约$10^{80}$个物质粒子与辐射。这些粒子的数目如此巨

大，就像一本页数奇多的书一样，带来的熵不容小视。这样一来，即使引力熵有所降低，新产生的这些粒子带来的熵也足以补偿这种降低。因而，总的熵，正如我们根据热力学第二定律预料的那样，实际上增加了。

但是很重要的一点是，暴涨膨胀虽然平滑了空间而且保证了一个各向同性、均匀、低熵的引力场，但同时也使得引力场实际贡献的熵与本应贡献的熵之间产生了一条巨大的鸿沟。虽然在暴涨过程中总的熵增加了，但却比本来应该增加的数量少得多。在这层意义上，我们可以说暴涨过程导致了一个低熵的宇宙：暴涨结束的时候，熵的确是增加了；但是增加的量并不是空间膨胀过程中的熵增量可比的。如果我们将熵比作财产税，那么暴涨过程就相当于纽约将撒哈拉沙漠纳入自己的版图：这样一来，总的税收自然是增加了，但是这种税收上的增加同面积上的增加完全不可比拟——面积惊人的撒哈拉沙漠只能贡献那么可怜的一点税收。

暴涨阶段一结束，万有引力就开始追讨那笔缺少的熵差。万有引力从空间不均匀性（当初量子涨落带来的小小不均匀性所埋下的种子）中聚集起来的每一块物质团——不管它是星系、恒星、行星，还是黑洞——都在增加着自身的熵，都在帮助万有引力实现它本该有的熵。考虑到这层意义，我们可以说，暴涨是这样一种机制：它制造了一个巨大的宇宙，但却只给了这个宇宙很少一点的引力熵，从而为接下来的几十亿年间的引力聚集做好了铺垫，我们今日见证的正是这一结果。所以，暴涨理论为时间之箭找到了方向：它首先为时间之箭安排了一个极低引力熵的过去；然后，时间之箭的未来就是熵增的方向。[4]

我们沿着第6章中讨论过的时间之箭所引导的方向走下去的话，就会遭遇第二个问题。从一个鸡蛋我们想到了生蛋的鸡，从生蛋的鸡想到了鸡吃的饲料，又从鸡吃的饲料想到了太阳的光和热，再从太阳的光和热想到了大爆炸中均匀分布的原初气体，就这样，我们随着宇

宙的演化回到了高度有序的过去，在旅程的每一个阶段我们都把低熵之谜留给了更古远的一个阶段。通过刚刚结束的讨论，我们认识到更早的暴涨膨胀阶段可以自然的解释大爆炸之后的光滑性和均匀性。但暴涨本身是怎么回事呢？我们能否解释这一串推敲的最初一环呢？我们最终能否解释为什么恰好存在暴涨膨胀发生的条件呢？

这个问题极为重要。不论暴涨理论能够解释多少谜题，只要暴涨理论没法发生，一切的讨论都是白费工夫。而且，由于我们无法回到过去直接看一看暴涨是否真的发生过，若想评判我们在为时间之箭设定方向方面是否取得了真正的进展，就要弄清楚达成暴涨膨胀所需的必要条件其可能性究竟有多少。换句话说，物理学家们对标准大爆炸理论依赖于精细调节成各向同性的初始条件非常不满，因为如此设置初始的动机仅来自于实验观测，而不是理论自身的要求。简单的为早期宇宙假定一个低熵状态令人难以满意；同样的，不经解释就直接将时间之箭安插到宇宙中也令人不安。乍看之下，暴涨理论取得了一些进展，使我们了解到标准大爆炸理论中假设的一些东西实际上来自于暴涨演化。可如果暴涨理论的启动也需要其他一些非常特别、极端低熵的条件，那我们岂不又回到原点了？我们只不过将大爆炸理论的特殊条件转换成了点燃暴涨的必要条件，而时间之箭的谜仍旧是个谜。

那么，暴涨发生的必要条件是什么呢？我们已经知道，暴涨子场的值只要在某段时间内在某个很小的区域内停留在势能碗的某一高地上，暴涨过程就将不可避免地发生。因此，我们的目标就是找出这种初始状态实际发生的可能性。如果最终发现这种初始条件很容易满足，我们就应该相信暴涨很可能实际发生过；但是，如果这种必要条件很难实现，我们就只好把有关时间之箭的问题再推迟一步作答——因为我们必须得先解释清楚启动一切的暴涨子场低熵状态是如何形成的。

在下面的讨论中，我将首先为大家讲讲在这个问题上已经理解得比较好的一面，然后再来谈谈有关这个问题人们还不清楚的地方。

## 玻尔兹曼的回归

我在前面的章节中曾经说过，我们最好将暴涨过程视作一次发生在已存在的宇宙中的事件，而不是将其视作一次创造宇宙的事件。虽然我们仍无法对前暴涨时期的宇宙形态给出一个无可争辩的说明，但还是让我们一起来看看在假定一切都处于完全平常的高熵状态下，我们能前进多少。明确起见，让我们将原初的前暴涨时期的空间想象成坑坑洼洼的样子；相应地，暴涨子场高度无序，其场值就像热锅中的青蛙一样跳来跳去。

如果你在一台没有作弊的角子机上耐心地玩下去，早晚有一天会转出3颗钻石；原初宇宙的高能量的狂暴状态迟早也会由于偶然的涨落而使小块空间中暴涨子场的值跳到正确、不变的位置上，从而诱发暴涨膨胀的爆发。如我们在前面的章节中解释的那样，计算表明，只需要有很小块的空间——只需要$10^{-26}$厘米见方——就能确保宇宙膨胀（紧随暴涨膨胀的标准大爆炸膨胀），将空间拉扯到比我们今日所见宇宙大很多的地步。因而，根本不需要预先假定或者简单断言早期宇宙的条件正好适合暴涨膨胀发生，按这种方式思考的话，只要有一块20磅重的东西有超微观涨落发生，周围只需是普通、平淡无奇的无序环境，就能产生暴涨所需的必要条件。

而且，正如角子机会转出大量不能赢的结果，在原初空间的其他区域上也有其他种类的暴涨涨落发生。大多数情况下，这些涨落不会是正好的值，也不会足够均匀以使暴涨膨胀得以发生（即使在$10^{-26}$厘米见方的区域中，不同位置处场值间的差异也将极其巨大）。但对我们来说重要的是，只要有那么一小块空间能够诱导暴涨发生就行，它将成为低熵链上最初的一环，并会最终将我们带到熟悉的宇宙。因为我们所看到的只有这么一个大宇宙，我们也就只需要宇宙这台大角子机成功一次即可。[5]

我们一路追寻宇宙的踪迹，发现了来自于原初混沌的统计涨落，

对时间之箭的这一解释与玻尔兹曼的原始想法有异曲同工之妙。回想第6章，玻尔兹曼认为我们今天看到的一切，来自于虽然不多但偶尔会有的总体无序中的涨落。而玻尔兹曼原始理论中的问题在于，偶然的涨落为什么会这么离谱，制造出了一个过度有序的宇宙，而本来不需要如此高度有序即可符合我们的生活所知。宇宙为什么会如此之大，存在着数以亿计的星系，而每个星系中又有数以亿计的恒星。宇宙为什么不走捷径呢？比如说，就制造出少数几个星系，甚至根本就只制造出一个星系。

从统计的观点看，较为温和的偶然涨落——带来一定程度的有序，但又不像我们今日看到的那么多——更是远远不可能。而且，因为平均说来熵是在增加的，所以玻尔兹曼的论证表明，我们所见的一切很有可能才刚刚得自于向着低熵态的偶然统计跃迁。回忆一下这是为什么：越靠近统计发生之时，所必须有的熵就越低（熵，一旦落到很低的位置，就要开始上涨，如图6.4所示。因而，如果涨落昨天发生，熵就落到昨天该有的位置，如果涨落10亿年前发生，熵就落到10亿年前该有的更低位置）。因而，时间越靠后，所需要的涨落就越不可思议、越不可能。所以，向着低熵态的跃迁很有可能才刚刚发生。但是，如果我们接受了这种结论，我们就没法再相信记忆、记录，甚至当前讨论背后的物理定律——而这是我们无法忍受的。

玻尔兹曼思想的暴涨化身的巨大优点在于，早期的一小点涨落——在极小块空间内，到所需条件的适度一跃——将不可避免地导致我们所熟知的巨大且有序的宇宙。暴涨膨胀一旦开始，那小块的空间将会被无情的拉扯到最少如我们今日所见宇宙的尺度上；因而，宇宙为什么没抄近路就毫不奇怪了；宇宙为什么如此之大，其中为什么有这么多的星系也就毫不奇怪了。从一开始，暴涨就给了宇宙一笔神奇的交易。低熵的小块空间被暴涨的杠杆作用放大到宇宙的巨大跨度。最为重要的是，暴涨带来的并不是什么古老的大宇宙，它带来的是我们的大宇宙——暴涨解释了空间的形状，解释了大尺度上的均匀性，

甚至还解释了“较小尺度”上的不均匀性，比如说星系以及背景辐射中的温度变化。暴涨利用涨落达到低熵状态，一下子解释了很多问题，并具有强大的预言能力。

玻尔兹曼可能一直都是对的。我们所看到的一切可能都来自于原初混沌那高度无序的状态的偶然涨落。在其思想的暴涨实现中，我们可以相信自己的记录，可以相信自己的记忆：涨落并不是刚才发生的。过去真的存在，我们的记录就是已发生事情的记录。暴涨膨胀放大了早期宇宙的那一小点有序——它一下子就把宇宙变得极为巨大，但同时却只有很小的引力熵——所以，接下来继续伸展以及星系、恒星和行星形成的那140亿年，就毫不足奇了。

事实上，这种方法告诉我们的还要更多。既然在百乐宫那众多的角子机上都有赢得大奖的可能，那么在高熵原初态和整体的混沌中，暴涨膨胀所需要的初始条件也没有理由非得来自于单独的一块空间。所以，安德烈·林德提出，可能分布在或这或那的很多小块空间都经历了平滑空间的暴涨膨胀。如果真是这样，那么我们的宇宙就只不过是偶然的涨落使物理条件刚好有利于暴涨发生（如图11.2所示）时所萌生的众多宇宙中的一个，或许还会有更多的宇宙继续产生出来。而因为其他这些宇宙将永远的与我们隔绝，所以我们可能永远都没有办法确认这种“多宇宙”图像是否正确。但是，仅作为一种理论框架的话，这种图像使得物理内涵既丰富又富于启发性。除了其他方面，它还为我们提出了另一种思考宇宙的方式：在第10章中，我将暴涨描述为标准大爆炸理论的“前端”，其中，所谓的爆炸是发生在瞬间的急速膨胀。但是，如果我们将图11.2中每一个宇宙的暴涨式发育视作其自身的爆炸，那么我们最好将暴涨视作包罗万象的宇宙学体系，在这个体系中，大爆炸就像演化一样，一个接一个的发生。因而，在这种方法中，我们并不是将暴涨纳入标准大爆炸理论中，而是将标准大爆炸理论纳入暴涨理论中。

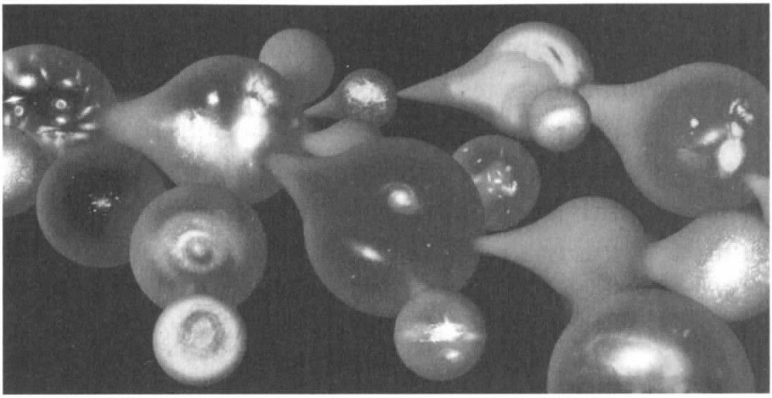


图11.2　暴涨可以反复发生，新的宇宙总是萌芽于旧的宇宙。

暴涨与鸡蛋

为什么你只能看到鸡蛋破碎而不会看到破碎的鸡蛋重新变得完整？我们所感受到的时间之箭来自何方？下面就是这一方法能够告诉我们的答案。通过普通的高熵原初态所常有的涨落中的某一瞬间机会，一块小小的20磅重的空间达到了能引发暴涨的条件。猛烈地向外膨胀将空间拉扯得极其巨大且极为平滑，随着暴涨趋近结束，暴涨子场将其因暴涨而被瞬间放大的能量以物质和辐射的形式均匀地填充到空间之中。随着暴涨子场的排斥性引力逐渐消失，普通的吸引性引力变得重要起来。而且，如我们所见，引力充分地利用了由量子涨落造成的微小不均匀性，使物质聚集成星系以及恒星，并最终形成了太阳、地球、太阳系中的其他天体，以及我们看到的宇宙所具有其他特点（如我们讨论过的那样，大爆炸之后差不多70亿年的时候，排斥性引力再次取得主导地位，不过这只与最大的宇宙尺度上的事情有关；对于较小的实体，比如单独的星系或者我们的太阳系没有直接影响。对于这些小的天体来说，引力只具有普通的吸引性）。太阳那相对低熵的能量，被地球上低熵的植物以及动物生命形式利用，产生出更为低熵的生命形式，这些低熵的生命形式通过热与消耗慢慢地引起总的熵增。最终，这一链条上产生出了母鸡、母鸡下蛋，剩下的故事你都

知道了：鸡蛋从你的厨房中的台子上滚落摔碎不过是宇宙不可抗拒的走向高熵状态过程中的小小一部分。暴涨拉伸使空间结构具有低熵、高度有序、均匀平滑的性质；这种性质类似于《战争与和平》中的页码按照正确的序号排列。正是这早期的有序状态——没有厉害的聚团或蜷曲，也没有庞大的黑洞——为宇宙接下来向着高熵方向的演化做好了准备，并带来了我们所感受到的时间之箭。以我们现在的理解水平，这就是所能获得的对时间之箭最完备的解释。

## 白璧微瑕

对我来说，暴涨宇宙学与时间之箭的故事非常有趣。从狂烈又活力十足的原初混沌中，产生出了均匀暴涨子场的超微观涨落，而这暴涨子场还不如乘飞机时所允许的手提行李重。它开启了暴涨膨胀，而暴涨膨胀又为时间之箭设定了方向，剩下的就都是历史了。

但是在讲这个故事的时候，我们做了一个未经证明的关键假设。为判断暴涨被启动的可能性，我们不得不指定暴涨没有出现之前的前暴涨时期的特征。我们预想的这一特别时期——狂乱、混沌、活跃——看起来很合理，但是将这种直观上的描述转换成数学语言被证明却极为困难。而且，这仅仅是猜测。在图10.3的模糊部分，没有这一信息我们就没法对暴涨启动的可能性做出令人信服的判断；而对这一可能性的任何计算都敏感的依赖于我们所作的假设。[6]

因为我们的理解中有这样的缺陷，所以我们能给出的最合理的结论是，暴涨提供了一个将看起来全无干系的问题——视界疑难、平坦性疑难、结构起源问题、早期宇宙低熵问题——捆绑起来的强大理论框架，并提供了一种对付所有这些问题的解决之道。这样感觉像是对的。但是再进一步，我们需要一个能对付得了模糊地带的极端特性——极端的热以及巨大的密度——的理论，那样我们才能对宇宙的最初时刻有一个清楚的、毫不含糊的认识。

我们在下面的章节中将会学到，这将要求有一个理论，这个理论必须能够跨越或许是过去80年间理论物理所面对的最大障碍：广义相对论与量子力学之间的鸿沟。很多科学家相信，一个较新的，所谓的超弦理论已经达成了这个目标。但是，如果超弦理论真是对的，那么宇宙的结构将比所有人能够想象出来的还要奇特很多。

# 第4部分
# 起源与统一

## 第12章　弦上的世界

弦论中的宇宙结构

让我们来想象这样一种宇宙，如果你想弄清这个宇宙中的任何一件事情，那么你必须首先完全弄明白关于这个宇宙的一切。在这个宇宙中，即使你只想稍稍了解一下行星为什么绕着恒星转，棒球为什么按着特别的轨迹飞，磁场或电池是怎么起作用的，光或者引力又是怎么一回事——总之是关于这一宇宙的任何一件事情——你都得先知道这个宇宙在最基本层面上的相互作用是怎样的以及这些相互作用是怎样作用到这个宇宙最基本的组成物质上才行。谢天谢地，我们的宇宙并不是这样。

如果我们的宇宙就是上面描述的那个样子，我们就没办法取得任何科学进步。几个世纪以来，科学之所以能取得长足进展，就是因为我们可以一点一滴的研究这个世界；每一个新的发现都使我们对这个世界的认识深入一步，我们就是这样一步一步揭开这个世界神秘的面纱。牛顿不需要任何原子方面的知识就可以在运动与引力的研究方面迈出一大步。麦克斯韦不需要知道电子和其他带电粒子方面的知识就可以推导出有关电磁场的强大理论。爱因斯坦在构建时空如何在引力场中弯曲的理论时也不需要先想清楚时空的原始形态。所有的这些发现，连同另外一些身为当代宇宙概念基础的伟大发现指引着人类不断前行；在这个前进过程中，那些人类暂时回答不了的基本问题总是被

堂而皇之的置于一边。即使没有人知道——即使现在也没人知道——给出所有这些谜题的究竟是一幅怎样的物理画卷，人们的每一个发现还是能对解释那些谜题贡献自己的一份力量。

换个角度我们可以发现，尽管今天的科学已经大大的不同于500年前的科学，但科学进步还是可以归结为新理论颠覆旧理论。更准确地说，新的理论总是在更加精确或更具有普遍性的框架下精练了旧有理论。牛顿的引力理论被爱因斯坦的理论超越，但是我们并不能因此就说牛顿的理论是错误的。当研究速度远远低于光速的物体运动以及强度不像黑洞附近那么强的引力场时，牛顿理论有着超乎想象的精确性。另一方面，这也并不是说爱因斯坦的理论只是牛顿理论的小小修正，爱因斯坦开启了一片全新的天地，在根本上改变我们关于空间和时间的概念。但是在牛顿理论的适用范围内（行星运动，人类日常生活中的运动问题），牛顿理论无可替代。

我们相信每一个新的理论都使我们更加接近事实的真相，但是是否有一个终极理论存在——一个再也无法改进的理论，因为它在可能的最深层面上为我们解释了宇宙的奥秘——则是一个没有人知道答案的问题。但即使这样，过去300年间的探索之路使人们有理由相信有一个这样的理论存在。宽泛地说，每一种新的突破，都是将更宽广范围内的物理现象归结到更少的理论庇护之下。牛顿的理论告诉我们使天体运行的力同使物体掉到地面上的力是同一种力。麦克斯韦的发现告诉我们电和磁只不过是同一硬币的两面。爱因斯坦的理论则告诉我们空间和时间是不可分割的，两者就像迈达斯那轻轻一点和金子的关系一样[56]。20世纪早期整整一代物理学家的理论发现告诉我们微观世界的种种神秘现象可以用量子力学精确地解释。晚近一些，格拉肖、萨拉姆和温伯格告诉我们电磁力和弱核力是同一种力——电弱力——的两种不同表现形式；而且，某些初步的间接证据表明强核力可能也是与电弱力统一在一起的。[1]从所有的这些中我们可以看出一种模式，那就是不断的由复杂到简单，从多样到统一。看起来，解释之箭

最终指向的将是一个强大的尚未被发现的理论，这一理论会将自然界中所有的力以及所有的物质统一到一个可以描述所有的物理现象的独一无二的理论框架下。

开启了现代统一理论之门的正是爱因斯坦，他穷尽30年的时光试图将电磁力与广义相对论统一到一个单独的理论中。在很长的一段岁月中，爱因斯坦独身一人苦苦寻觅着统一理论，但是他的热情却使他离开了物理学家群体的主流。在过去的20年间，寻求统一理论之梦再度燃起；爱因斯坦孤独的寻梦之旅已经成了一代物理学家的驱动力。不过相比于爱因斯坦时代，问题的焦点已经有所变化。尽管我们还没有一个可以将强核力与电弱力统一起来的完美理论，但是我们已经可以用基于量子力学的统一语言描述这3种力（电磁力、弱力和强力）。但是广义相对论描述第4种力所用的语言，仍然游离于理论框架之外。广义相对论是一个经典理论：没有使用任何的量子力学概率概念。现代统一计划最初的一个目标就是将广义相对论与量子力学统一起来，然后在同样的量子力学框架下描述所有的4种力。而人们已经发现这可能是理论物理学家所遇到的最难对付的一个问题。

现在我们一起来看看这究竟是为什么。

量子涨落与真空

如果要我来选出量子力学最特别的性质，那么我将选出不确定原理。诚然，概率与波函数提出了全新的理论框架，但是真正将量子力学与经典物理区别开的却是不确定原理。还记得吗？17世纪、18世纪时的科学家们相信，对物理实体的完备描述可归结为搞清楚构成宇宙的全体物质的位置与速度。随着场的概念在19世纪出现，这一观念被应用于电磁场和引力场，于是转而变为搞清楚在空间中的任意位置处每种场的值——就是每种场强——以及每种场的值的变化率。但是到了20世纪30年代，不确定原理改变了这种观念，它告诉人们，我们根本没办法同时搞清楚一个粒子的位置和速度；我们也没办法同时知道

空间中某一位置的场强及其变化率。量子力学的不确定原理不允许我们同时知道。

正如我们在上一章中讨论的那样，不确定原理使微观世界成为动荡的王国。在更前面的章节中，我们曾讨论过由于不确定性导致的暴涨子场量子涨落；而不确定原理可以应用于所有的场。电磁场、强核力和弱核力场以及引力场，都可以归结为微观尺度上狂暴的量子涨落。事实上，这些场的涨落甚至在一般认为的既没有物质也没有场的真空中也同样存在。这一观念极其重要，不过要是你之前没有接触过这些问题的话，可能会感到非常困惑。如果空间中的某一区域什么也没有——也就是说它是真空——那还有什么东西可以涨落呢？好吧，想一想，我们已经知晓了什么都没有这种说法是非常微妙的，现代理论中的希格斯海就存在于整个空间。我现在所说的量子涨落就是要使什么都没有这个概念变得更加微妙。下面就是我要讲的真正意思。

在量子力学诞生之前（以及希格斯物理学诞生之前）的物理学中，如果某一空间区域中没有粒子并且每种场的值都为零，那么我们就说这一空间区域是完全空的。[57][2]现在我们加上不确定原理再来看看这一经典概念——空。如果一个场的场强为零，那我们就既知道这个场的场强——零，也知道这个场的场强变化率——也是零。但是根据不确定原理，我们没办法同时知道这两个值的大小。如果一个场在某一时刻具有确定大小的场强，目前我们说它为零，不确定原理就会告诉我们其场强的变化率将是完全随机的。完全随机的变化率意味着场强在接下来的时刻会随机涨高落低，即使在我们通常认为完全空荡的空间中也是如此。所以“空”在直觉上的概念——所有场的值都为零——与量子力学是完全不相容的。一个场的值可以在零的上下涨落，但却不能在一段时间内在空间中的某一区域中始终保持为零。[3]如果用专业的术语来说，物理学家们会将其形容为场具有真空涨落。

真空场涨落的随机性保证了在所有的微观区域上，既有涨高也有落低，因而其平均为零。这一现象就像是大理石的表面：虽然用肉眼

看起来光滑如镜，但是如果用电子显微镜观察一下微小尺度上的大理石表面，我们就会发现其实是参差不齐的。但是，虽然我们不能直接看到那些真空涨落，半个世纪前的人们仍然想到了一些虽然简单但却实用的方法，肯定了量子涨落（即使在真空中）的实在性。

1948年，荷兰物理学家亨德利克·卡西米尔发现了实验测量电磁场真空涨落的方法。根据量子力学，电磁场在真空中的涨落可以呈现出一系列的波纹，如图12.1（a）所示。卡西米尔首先想到，如果在真空中放置两块普通的铁板，如图12.1（b）所示，那么真空中的涨落形状就会有所改变。即根据量子力学方程，铁板之间区域的量子涨落要稍稍弱于其外区域的量子涨落（仅当电磁场涨落在铁板处的值为零时成立）。卡西米尔仔细分析了场涨落的减小所带来的效应，发现了一些非常特别的东西。正如某一区域的空气减少会导致压强的不平衡（例如，在高海拔的区域，空气稀薄，因而你的耳膜所感受到的压强就会小些），铁板之间量子涨落的减小也会导致电压的不平衡：两块铁板之间的量子涨落变得比铁板之外区域的量子涨落小的话，所导致的电压差会使得两块铁板彼此接近。

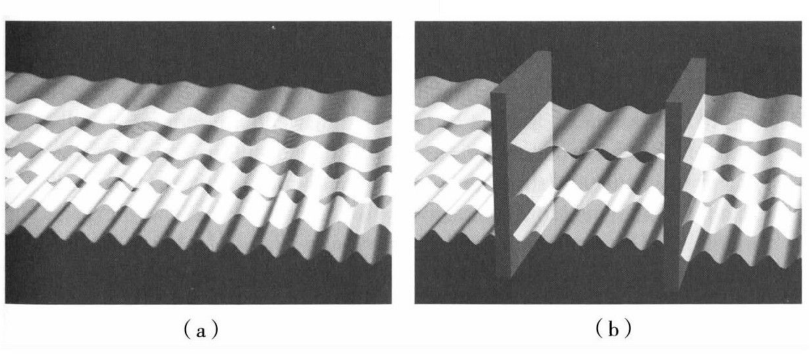


图12.1　（a）：电磁场的真空涨落。（b）：两块铁片之间以及其外的真空涨落。

想想看吧，这有多么奇怪。你就仅仅把两块平常得不能再平常的铁板彼此平行地放到真空中。而这两块铁板的质量又非常之小，以至于它们之间的引力相互作用完全可以忽略。周围再也没有其他的什么

东西了，于是你想当然的会认为这两块铁板就会那样静静的待着。但是卡西米尔的计算却说事情并不是这样。他的计算告诉我们这样的两块铁板会由于真空涨落造成的鬼魅般的压力而彼此靠近。

在卡西米尔提出他的这些论断之初，实验设备还没有精良到足以完成这种实验的地步。10年之后，另一位荷兰物理学家马库斯·斯巴尼开始尝试实验检验卡西米尔力。从那以后，人们又进行了大量的精确实验。比如1997年，其时在华盛顿大学的史蒂夫·拉莫雷奥克斯在5%的精度上确证了卡西米尔力[4]（两块扑克牌大小的铁板如果间距为万分之一厘米，其间的卡西米尔力就相当于一滴眼泪的重量；由此可见，测量卡西米尔力是一件多么难的工作）。现在的科学家们几乎不再怀疑直觉上的真空概念——静止、冰冷、空无一物的空间——大错特错了。由于量子力学的不确定性，真空中有着丰富的量子行为。

20世纪的科学家们花了很多力气来发展用以描述电磁力、强核力与弱核力的量子行为的数学工具。这些力气并没有白费：用这些数学工具理论计算出来的结果可以在非常高的精度上与实验上测得的结果相比较（比如，对电子磁性质的量子效应的理论计算与实验结果的符合程度就高达十万分之一的精度）。[5]

但是，物理学家们几十年来一直都很清楚，在这些成就之外，量子涨落与物理定律之间有很多不和谐之处。

涨落与不谐[6]

目前为止，我们还仅限于讨论空间中的场的量子涨落。那么空间本身的量子涨落呢？虽然听起来可能有点奇怪，不过这只是量子场涨落的另一个例子——可这个例子着实棘手。在广义相对论中，爱因斯坦提出引力可以用空间的蜷曲和弯曲加以描述；这位伟人证明了引力场可以通过空间（更具普遍性的说法是时空）的形状或几何来展现自己。就像其他的场一样，引力场也可以归结为量子涨落：不确定原理

保证了在小尺度上，引力场也可以上下波动。既然引力场与空间的形状是同一个意思，那么引力场的这种涨落也就相当于空间本身的涨落。就像不确定原理的众多例子一样，在人类日常生活的距离尺度上，空间的涨落太小以至于没法为人们所直接感知，我们周围的一切还是光滑、宁静，尽在掌握。但是随着所观测尺度的减小，不确定性就会增大，量子涨落也变得越来越明显。

图12.2所示的就是这一情形，我们把空间逐渐放大以发现更小尺度上的空间结构。图中最底层示意的是平常尺度上空间的量子涨落，正如图所示，我们什么都看不到——量子涨落太小以至于无法观测，空间还是宁静平坦。我们进一步放大观测区域，就会观测到一定程度的涨落。在图的最上层，空间结构的尺度已经比普朗克长度——1厘米的十亿亿亿亿分之一（$10^{-33}$）——还要小，这时的空间变得沸腾躁动，喧嚣不已。从图中我们可以清楚地看出，平常所谓的那些“左右”、“前后”、“上下”等概念在小尺度的狂乱中全部失去了意义。还不只这些，考虑那些小于普朗克时间——1秒的千亿亿亿亿亿分之一（$10^{-43}$秒，在这一时间间隔内光可以走普朗克长度那么远）——的时间尺度时，我们平常的“以前”、“以后”这样的时间概念也都失去了意义。就像一张模糊不清的照片，图12.2中的波动使得我们不可能分辨小于普朗克时间的两个时刻。这一切的结果就是，在小于普朗克距离与普朗克时间的尺度上，量子不确定性使得宇宙的结构扭曲混沌，通常的空间和时间的概念不再具有任何意义。

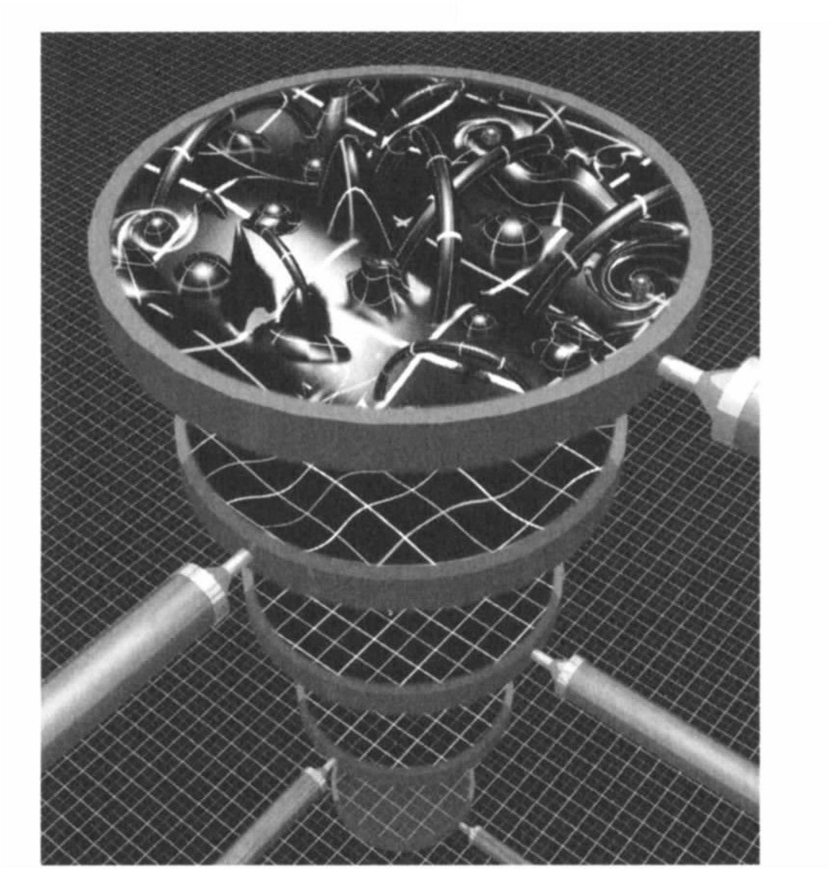


图12.2　将空间连续放大后我们发现，普朗克尺度之下的空间由于量子涨落而躁动不安（图中所示的是想象中的放大镜，每一个可以放大1000万倍到1亿倍）。

虽然细节上非常古怪，但是图12.2告诉我们的无非就是一个我们已经非常熟悉的事实：与某一个尺度有关的概念和结论无法应用到其他的尺度上。这是物理学的一个重要原理，我们一遍又一遍地遇到这个原理，即使在那些非常普通的知识中也能遇到。以一杯水为例，从日常生活的尺度来看，这杯水不过是光滑均匀的液体；但是我们在微观尺度上来看的话就不再是这样。小尺度上，光滑的图像被另一种完全不同的景象代替，那就是彼此间距很大的分子和原子。类似地，图12.2告诉我们的是，爱因斯坦的那些平滑弯曲几何式的空间和时间的

概念，虽然可以在大尺度上强有力且高度精确的描述宇宙，但在极小的长度和时间上就不再有效了。物理学家们相信，就像那杯普通的水一样，空间和时间光滑的形象只能是一种理论近似，在超小尺度上，这种近似必将让位于更加基本的理论框架。而这一理论框架究竟是什么——时间和空间的“原子”和“分子”究竟是什么——则是物理学家们以极大的热情苦苦追寻的问题。不过物理学家们还没能找到答案。

即使我们还没有最终答案，图12.2仍然清楚地告诉我们：小尺度上，广义相对论所带给空间和时间的光滑形象必将被量子力学带来的狂躁涨落的形象替代。爱因斯坦广义相对论的核心原理——空间和时间形成柔和弯曲的几何形状，与量子力学的核心原理——不确定原理，这一原理告诉我们最小尺度上的时间和空间狂野动荡——之间存在着激烈的冲突。广义相对论与量子力学在核心层面上的这种冲突使得调和这两个理论成了过去80年间物理学家面临的最大困难。

这重要吗

实际上，广义相对论与量子力学的不相容性总是通过一种特别的方式展现自己。如果你将广义相对论与量子力学的方程组合到一起，那么你总会遇到一个麻烦：无限大。这是一个大问题，因为无限大毫无意义。实验学家们从未测到过任何无限大的数，刻度针从不曾指向过无限大，仪表永远也不会达到无限大，计算器处理不了无限大，一个无限大的结果差不多总是毫无意义。所有这一切告诉我们的就是：当把广义相对论的方程和量子力学的方程组合到一起的时候，出了什么大毛病。

需要注意的是，这里的问题并不同于我们在第4章中讨论量子非定域性时提过的狭义相对论与量子力学之间的问题。在第4章中我们了解到，为了将狭义相对论的原则（特别是所有匀速运动的观测者之间的对称性）与纠缠粒子的行为协调一致，我们需要对量子测量问题有一个更加完备的理解（详见第4章“纠缠与狭义相对论：反方观

点”小节）。这一未被完全解决的问题并没有带来数学上的不自洽或是方程结果的无意义。恰恰相反，将量子力学与狭义相对论结合起来的方程给出了科学史上最精确的理论预言。狭义相对论与量子力学之间的小小麻烦告诉人们的是有一个研究领域需要进一步探索，而并不影响将两个方程结合起来的理论预言能力。而广义相对论与量子力学的不相容却使得理论预言的能力完全丧失了。

不过即使这样，你仍然可以提出这样的问题：广义相对论与量子力学之间的不相容性真的有什么要紧吗？没错，将两个方程组合起来确实会带来无限大，不过你真的需要将它们组合起来吗？几十年的天文学观测已经证实，广义相对论可以以难以企及的精确性描述恒星、星系甚至整个宇宙的扩张这些宏观世界的物理；大量的实验同样证实量子力学在描述分子、原子、亚原子粒子这些微观世界的物理时同样威力强劲。既然这两个理论在其各自的领域内运转良好，我们为什么非要将它们组合起来呢？就让它们一直分开不是很好吗？为什么不就用广义相对论讨论那些又重又大的家伙，用量子力学讨论那些又小又轻的家伙呢？这样我们就可以庆贺人类已经在如此宽广的领域上了解了这个世界的物理现象。

实际上，这正是20世纪早期以来大多数物理学家一直做的事情，毫无疑问，这一直都是一种能获得丰富成果的好方法。在两种不同的理论框架下，物理学家们成就斐然。不过，仍然有很多理由要求广义相对论和量子力学之间的对抗必须得以调和。下面我们就来谈谈其中的两个理由。

首先，在大统一理论的层面上来看，人们很难想象统治我们这个宇宙的基本原理由两个彼此不能相容的理论组成。我们很难想象宇宙会把一切的事物泾渭分明的划分为两派，一派由量子力学描述，另一派则由广义相对论描述。把宇宙划分为两个不同派别的办法看上去是一个纯粹人为的办法，而且还非常笨拙。很多人相信，一定会有一个真正的深层次的统一理论将广义相对论与量子力学的矛盾调和起来，

这样的一个理论可以应用到一切尺度上的物理。我们只有一个宇宙，因而很多人相信，我们应该也只有一个理论。

另外，尽管大部分的事物要么又大又重，要么又小又轻，因而从实践的角度看，可以利用广义相对论或者量子力学分别加以描述。不过，这个不是绝对的。黑洞就是一个很好的特例。根据广义相对论，组成黑洞的所有物质都被挤压到黑洞中心的一个很小的点上。[7]这就使得黑洞的中心既极其的重又极其的小，因而必须依靠被分开的两个理论：我们需要广义相对论，因为黑洞的大质量会产生一个充实的引力场；我们也需要量子力学，因为所有的质量都被挤压到一个很小的尺度上了。但我们一旦将广义相对论和量子力学的两个方程组合起来，这个方程就会垮掉，所以没有人能够计算出黑洞的中心会发生什么。

黑洞就是一个好例子。不过如果你是一个真正的怀疑论者，那么你或许会问：这是不是也是一个我们不需要考虑的问题呢？因为我们如果不跳到黑洞的里面就没办法看到黑洞的里面发生了什么；而我们要是跳进去了，我们又不能将黑洞里面的情况报告给黑洞外面的世界，因而我们并不需要为黑洞里面是什么情况这样的问题而烦恼。但是对于物理学家来说，要是存在现有物理定律垮掉的领域——不管这一领域看起来多么古怪，那这就是一个真正的危险信号。只要已知的物理定律在某些情形下垮掉，那就明确地意味着我们还没有真正掌握最深层次的物理。毕竟宇宙总是正常运行，宇宙并没有垮掉。关于宇宙的正确理论至少应当满足这一标准。

好吧，这很合理，不是吗？但在我看来，由于量子力学与广义相对论的冲突而带来的问题中有一个更加需要尽快加以解决。我们再回头看看图10.6。可以看到，在将宇宙的演化串成一线方面，我们已经迈出了一大步，各个时期的演化前后一致且具有可预言能力。但是事情还没有最终完结，因为我们还没有彻底搞清楚接近宇宙诞生的时期所发生的事情。最初的时刻还是具有令人迷惑不解的神秘，那就是时

间、空间的起源以及基本性质。那么是什么使我们不能揭开最初时刻的神秘面纱？就是量子力学与广义相对论之间的冲突。大质量的定律与小尺度的定律之间的矛盾使得我们没法补全宇宙演化模糊不清的那部分，宇宙形成之初的物理我们还是没办法洞察。

要理解这一点，让我们像在第10章中那样，倒过来放映一下宇宙演化这部片子，从膨胀的宇宙往回想象大爆炸。反过来想的话，每一种分散开来的东西又聚合到一起，我们的电影继续回放，宇宙变得越来越小，越来越热，越来越密。我们越接近时间上的零，整个可观测的宇宙也会变得越来越小，先是小到太阳那么大，接着只有地球那么大，然后只有保龄球那么大，梨那么大，一粒沙子了——电影不断回放，宇宙越变越小。终于在某个时刻，宇宙只有普朗克长度那么大——1厘米的十亿亿亿亿分之一，而这个尺度上的量子力学和广义相对论又开始闹矛盾了。此刻，产生现今可观测宇宙的所有质量和能量都被包纳在一个小于原子大小万亿亿分之一的小点内。[8]

如同黑洞中心的情况一样，早期宇宙的研究也需要求助于不相协调的两个理论：早期宇宙的大密度需要使用广义相对论来研究。而早期宇宙的超小尺寸又要求使用量子力学。于是，将两个方程组合到一起，一切又变得糟糕了。放映机卡住了，我们关于宇宙的回放只能到此为止了，于是我们还是不知道宇宙最初的那一刻。由于广义相对论和量子力学的冲突，我们仍然对早期宇宙一无所知，图10.6的开端还是只能混沌一片。

如果我们想要搞清楚宇宙的起源——所有科学中最深层次的一个问题——我们就必须解决广义相对论与量子力学之间的冲突。我们必须攻克由于“大”的定律与“小”的定律之间的矛盾而带来的问题，将两者融合成和谐一致的理论。

看似不可能的解决方式[58]

正如在爱因斯坦和牛顿身上所展现出来的那样，科学上的重大突破有的时候纯粹是来自于某个科学家令人意想不到的天才。不过这样的时候并不多见。更多的时候，科学突破是由多位科学家的集体智慧催生的，每一个都在别人的基础之上做出进一步的工作，集腋成裘，最后取得一位科学家难以企及的成就。某位科学家想到的点子可能会促使其同事发现一些以前人们未曾注意到的关系，而这些新发现的关系可能会引发一次重要的突破，于是又开始了新一轮的科学发现。宽阔的眼界，熟练的技巧，灵活的头脑，对未曾预料到的联系的接纳能力，勤奋的工作，以及难以想象的运气都是科学发现的关键要素。近些年来，没有什么理论比超弦理论的发展更适合展现这一点。

很多科学家相信超弦理论将成功地调和量子力学与广义相对论。我们会看到，有理由相信超弦理论带给我们的将不止这些。尽管超弦理论目前还在研究中，但它很有可能是一个能够统一所有的力与所有的物质的理论，超弦理论很有可能实现甚至超越爱因斯坦之梦。我，还有很多科学家都相信，目前的研究仅仅是一个绚烂的开始，超弦理论最终将带给我们关于宇宙的最基本定律。然而，超弦理论并非孕育于某个试图达到这些伟大的长远目标的天才方法中。恰恰相反，超弦理论的历史中有的是偶然的发现，错误的开始，误失的良机，以及几乎被终结的命运。更确切地说，超弦理论是为了解决错误的问题而做出的正确发现。

1968年，加布里埃尔·维尼齐亚诺还是CERN的一位年轻的博士后研究员。和当时的许多物理学家一样，他致力于通过研究世界范围内各种原子对撞机上高能粒子的对撞结果来探索强核力。对数据中具有的模式和规律性经过数月的分析研究后，维尼齐亚诺神奇地发现这些数据同某一深奥的数学领域有着令人意想不到的联系。他发现有关强核力的这些数据同著名的瑞士数学家利昂纳德·欧拉在200多年前发现的一个公式（欧拉贝塔函数）可以精确匹配。也许这听起来没什么特别的——物理学家们总是使用不可思议的公式来研究问题——但在这

里却着实是一个带有超前意味的意外发现，就像马车和缰绳跑到了马的前面一样。虽然并不总是，但大部分时候，物理学家都是先对所研究的问题有一个直观的物理图像，充分理解了他们正在探讨的物理问题之下掩盖的基本原理之后，物理学家们才寻求正确的方程来给他们的直观物理图像建立一个坚实严格的数学基础。维尼齐亚诺则不是这样，他直接就得到了方程；维尼齐亚诺的天才之处在于从纷繁复杂的数据中发现了特别的规律性，并将这一规律性同200年前纯粹来自于数学的公式联系起来了。

不过，维尼齐亚诺虽然得到了公式，但他却不知道如何解释这一公式为什么会有效。为什么欧拉贝塔函数会和影响粒子的强核力有关？维尼齐亚诺没有想清楚其中的物理图像。接下来的两年情况仍未改观。直到1970年，斯坦福的莱昂纳德·萨斯金、尼尔斯·玻尔研究所的霍奇·尼尔森，芝加哥大学的南部阳一郎等人才分别弄清了维尼齐亚诺发现的物理基础。这些物理学家证明，如果将两个粒子之间的强核力用一根连接粒子的极其细小的如橡胶管一样的绳子来解释的话，那么维尼齐亚诺和其他人所共同关注的量子过程就可以用欧拉公式描述。这些很小的弹性绳子就是所谓的“弦”。终于，马又跑到了马车的前面，弦论正式诞生了。

但先别忙庆祝。对于那些参与了这次研究的人来说，想清楚维尼齐亚诺公式的起源实在非常的有成就感，因为那表明物理学家们正在一步步解开强相互作用的神秘面纱。不过，这一发现并未掀起普遍性的狂热情绪，而且还差得很远。事实上，萨斯金的论文甚至遭到了期刊编辑部的退稿，理由是这一工作毫无意趣。萨斯金曾回忆那段经历：“我很吃惊，深受打击，非常沮丧，只好回家借酒浇愁。”[9]尽管最后他和其他人有关弦的论文都被发表出来了，但是立即又遭受了两次毁灭性的挫折。仔细研究20世纪70年代早期的大量有关强核力的实验数据后，人们发现弦论的方法并不能非常精确的符合最新发现的结果。接下来，量子色动力学（QCD）出现了，这一基于传统的粒子和

场——而不是弦——的理论可以令人信服地解释所有的实验数据。所以到了1974年，至少乍看起来，弦论遭到了重大的打击。

约翰·施瓦茨是弦论最早的狂热者之一。他曾经告诉我，从一开始他就觉得弦论深刻而意义重大。施瓦茨花费了数年的时间用以研究弦论方方面面的数学问题；抛开其他的成果不提，这一系列的研究导致了超弦理论——我们将会看到，超弦理论是原始弦论的一个重要的升级版——的发现。但是随着量子色动力学的巨大成功及在弦论框架下描述强核力的失败，在弦论上继续走下去似乎已无必要。不过，施瓦茨并没有放过弦论和强核力的不相匹，他不允许自己略掉这个问题。弦论的量子力学方程预言了一个非常特殊的粒子，这个粒子可以通过原子对撞机上的高能粒子碰撞大量的产生出来。就像光子一样，这一粒子的质量为零，但其自旋却为2。粗浅地说，这意味着这个粒子比光子转的快2倍。没有任何实验曾经发现过这样的一个粒子，因而这个粒子仅仅是弦论的众多未被证实的预言中的一个。

施瓦茨和他的合作者乔·谢尔克完全搞不清楚这个莫明其妙的粒子。直到某一天，他们将这个粒子与另外一个完全不同的问题联系了起来，这才取得了实质性的突破。尽管没有人能将广义相对论和量子力学结合起来，物理学家们还是可以定出一个成功的统一理论应有的一些性质。我们在第9章中曾经说过，微观层面上，电磁场通过交换光子来传递电磁力；引力场也是如此，只不过引力场交换的是另外一种粒子——引力子（基本粒子，引力的量子束）。虽然实验上尚未发现引力子，但是理论分析告诉我们引力子至少要有两个性质：无质量；自旋为2。引力子的这两个性质启发了施瓦茨和谢尔克——引力子的这两个性质正是弦论预言的那个讨厌粒子所具有的性质——促使他们迈出了大胆的一步；于是，看似失败了的弦论取得了梦幻般的成功。

施瓦茨和谢尔克提出，弦论根本就不应当被看做是强核力的量子理论。他们认为，虽然弦论是在探索强核力的过程中发现的，但这个

理论实际上是另一个完全不同的问题的答案。弦论实际上是第一个引力的量子理论。施瓦茨和谢尔克宣称，弦论所预言的自旋为2的无质量粒子正是引力子，而弦论的方程是引力的量子力学描述的具体表示。

施瓦茨和谢尔克于1974年发表了他们的论文。两人本希望这一设想会引起物理学家的广泛重视。但事与愿违，没有什么人对他们的理论感兴趣。现在回头来看，我们完全可以明白这是为什么。他们的想法似乎是非得为弦论找点什么用武之地。在解释强核力失败后，弦论的支持者们似乎不肯接受失败，他们好像竭尽全力也要找个能用得上弦论的地方。而且，在施瓦茨和谢尔克的理论中，弦的尺寸必须极大的改变一下，以便弦论中的候选引力子可以提供人们熟知的引力强度。我们都知道引力非常之弱[59]，而且根据弦论，越长的弦所传递的引力也会越强。基于这样的原因，施瓦茨和谢尔克发现他们的弦必须极其小才能够传递像引力那么弱的力；这样的弦必须小到普朗克长度，比之前作为强核力的理论时小了万亿亿倍。这样的情况无异于火上浇油。怀疑者们尖锐地指出，没有任何实验仪器有可能看到这么小的弦，这也就意味着这个理论完全不能用实验检验。[10]

另一方面，更加传统的非弦论式的点粒子与场的理论在20世纪70年代取得了令人目不暇接的成就。理论学家的大脑，实验学家的双手，全都被一个又一个实实在在的问题占据；人们不停地探索着新的理论，不断地用实验检验着理论的预言。既然在一个已经经受住了实践检验的框架下有这么多激动人心的工作等着人们去做，人们为什么要转投弦论呢？在这种情绪的感染下，尽管物理学家们知道他们的传统方法在调和广义相对论和量子力学方面存在着重大的问题，但却没把这个问题当成一个亟待解决的问题。所有的人都承认这是一个重大的问题，未来的某一天我们必须面对这个问题。但是，在丰富的非引力工作的诱惑下，量子化引力这个难题还是扔到一边留着以后再说吧。最后，还有一点要知道的是，弦论在20世纪70年代中期还远远未

形成体系。有一个引力子的候选者当然是一个成功之处，但是更多的概念性或技术性问题都还没有解决。弦论看起来很难克服那些未解决的困难问题，这个时候加入弦论的研究中多少带有一定的冒险意味。谁知道什么时候，弦论可能突然就死掉了。

但施瓦茨仍旧态度坚决。他相信弦论——第一个看似可能的用量子力学的语言描述引力的方法——的发现必定是一个重大的突破。如果大家都不感兴趣，好吧，没关系。反正他自己决意跟进，继续探索这个理论。人们真正注意到这个理论的时候，弦论已经发展到一定程度了。施瓦茨的果断具有真正的预见性。

20世纪70年代后期至80年代早期，施瓦茨与当时在伦敦玛丽女王学院的麦克尔·格林合作，一道解决弦论面临的一些技术性障碍。首要的问题是所谓的反常。我们不在这里讨论这个问题的细节。简单地说，反常是一个很恶劣的量子效应，它可以通过破坏某些不可撼动的守恒律——比如能量守恒——来毁掉一个量子理论。一个可行的量子理论必须没有反常。早期的研究发现弦论中具有反常，反常的出现是弦论没能引起人们兴趣的一个主要的技术性原因。即使引力子可以使弦论成为一个引力的量子力学理论，反常也会使得弦论遭受来自于其自身的数学不自洽的困扰。

但是，施瓦茨认识到问题并没有坏到毫无办法的地步。或许，完整的计算后，人们会发现各种量子贡献带来的反常会在正确的组合之后彼此相消。于是，格林和施瓦茨承担了计算这些反常的艰苦工作。两人在1984年的夏天终于挖到了真正的宝藏。一个暴风雨夜，在科罗拉多阿斯本的物理中心工作到很晚的格林和施瓦茨完成了这一领域最重要的计算。计算结果表明，所有可能的反常以一种神奇的方式的确彼此相消了。他们发现，弦论中并没有反常，因而也无须遭受数学不自洽的困扰。格林和施瓦茨令人信服地证明了弦论在数学上是可行的。

这一次，物理学家们终于认真听他们的报告了。20世纪80年代中期，物理学的气候明显发生了变化。除引力之外的3种力的很多重要性质都已经理论算出且经过了实验检验。尽管还有很多重要细节尚未解决——直到今天也没解决——物理学家们已经开始着手对付另一个重大难题：如何将广义相对论与量子力学合并起来。这时，格林和施瓦茨走出了不被注意的物理学小角落，带着明确的、数学上自洽的、美学上也受欢迎的弦论猛地出现在公众面前，来告诉人们如何解决广义相对论和量子力学的合并问题。几乎一夜之间，弦论的研究者从最初的两人变成了上千人。第一次超弦革命到来了。

第一次革命

我于1984年秋天在牛津大学开始了我的研究生学习。接连好几个月，走廊里到处都是谈论第一次超弦革命的嗡嗡声。那个时候互联网还不发达，各种传闻还是快速散播有关信息的主要渠道。每天都能听到新突破的消息。研究人员普遍认为自从量子力学诞生的最初岁月以来，物理学界的气氛还未曾如此躁动。甚至有人严肃地谈论着理论物理的尽头近在咫尺。

对于大家来说，弦论还是新事物。早期的时候，弦论的细节还不能算是常识。我们这些在牛津的人非常幸运：麦克尔·格林那个时候曾专门到牛津作过弦论方面的报告，我们大多数人都开始了解弦论的一些基本思想以及重要主张。弦论所宣称的内容令人印象深刻。简单地说，弦论说了以下几点：

以一片任意事物为例——可以是一块冰，一块石头，一张铁片——我们想象着将它一分为二，然后再一分为二，一直这样做下去。我们一直切到非常小的尺度上。大约2500年前，古希腊人就提出了按这样的过程追寻最细微、不可再切、不可分割的成分的问题。现在我们已经知道这样做早晚会遇到原子，而原子并不是古希腊人要的答案，因为原子还能够被切成更细的组分。原子是可以切开的。我们已

经知道，原子是由原子核和云集核外的电子组成；而原子核又是由质子和中子组成。20世纪60年代末，斯坦福直线加速器上的实验发现中子和质子也是由更基本的物质组成：每一个质子和中子都是由3个称为夸克的粒子组成。我们在第9章曾提到过这些内容，也可以参看图12.3（a）。

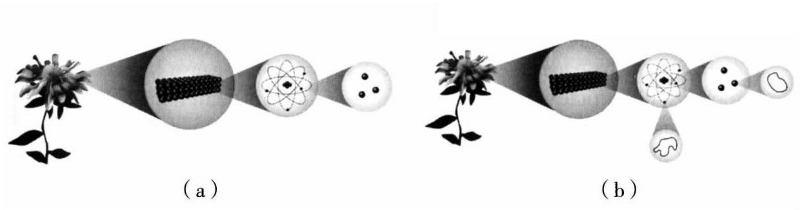



图12.3 （a）：传统理论将电子和夸克视为物质的基本组成。（b）：弦论则将每一个粒子看成振动的弦。

在由高度精确的实验支持的传统理论中，电子和夸克被视为无空间结构的点粒子；如果按这种方式看的话，电子和夸克就代表着尽头——在物质的微观结构中能发现的大自然的最后一个俄罗斯套娃。而在这里弦论要登场了，它要挑战传统理论。以弦论的观点看，电子和夸克并不是没有尺寸的粒子。传统的点粒子模型只不过是一种近似，每个粒子真正的样子是细小振动着的能量丝，我们将其称之为弦，如图12.3（b）所示。这些振动能量的线没有厚度，只有长度，因而弦是一维的实体。可是，弦实在太小了，比一个单个原子核还要小万亿亿倍（$10^{-33}$厘米）。所以，即使用我们最高级的原子对撞机来观测弦，我们看到的也只可能是点。

因为我们对弦论的理解还远未完备，所以没有人知道弦论是否就是故事的尾声——如果弦论是正确的，那么它就是最后一个俄罗斯套娃吗？弦是否也是由更基本的成分组成的呢？我们稍后再回到这个问题上，现在姑且按照历史发展，假定弦论就是一切的终点；我们就将弦先看做是宇宙最基本的结构。

弦论与统一

刚刚简要介绍了一下弦论，为了更好地展示弦论的强大之处，我有必要更加完整的讲一讲传统的粒子物理。过去的几百年，物理学家们一路磕磕绊绊的追寻着宇宙的最基本结构。人们发现，差不多世上所有的一切都是由前面提到的夸克和电子——如第9章中所述，更准确的说法是电子和两种夸克，质量和电荷分别不同的上夸克和下夸克——组成的。而实验告诉我们，宇宙中还存在着其他更加古怪的粒子种类，这些粒子并不出现在我们平常见到的事物中。除了上夸克和下夸克，实验上还发现了另外4种夸克（粲夸克，奇异夸克，底夸克和顶夸克）和另外两种很像电子却要重一些的粒子（μ子和τ子）。大爆炸之后很有可能存在很多这些粒子，但是到了今天，人们只能在高能对撞机上看到它们的身影了。除此之外，实验上还发现了3种幽灵般的粒子，即所谓的中微子（电子中微子，μ子中微子和τ子中微子）。中微子在铅中穿行万亿千米就像我们在空气中行走一样自如。所有的这些粒子——电子和它的弟兄，6种夸克和3种中微子——就是现代物理学家对古希腊的最小物质组成问题的答案。[11]

所有的这些粒子可以分为三“代”，如表12.1所示。每一代包括两个夸克、一个中微子和一个相应的电子类的粒子；不同代中相对应的粒子的区别只是质量不同。按代分类虽然使得粒子的种类看起来有规律可循了，但是这种粒子还是能搞得你有点头晕（甚至是眼花缭乱）。不过别怕，弦论的好处现在就体现出来了。弦论最美妙的一点就是能用一种方法驾驭这种明显的复杂性。

表 12.1 三代基本粒子及其质量（与质子质量对比）

| 一代 | | 二代 | | 三代 | |
|---|---|---|---|---|---|
| 粒子 | 质量 | 粒子 | 质量 | 粒子 | 质量 |
| 电子 | 0.00054 | μ子 | 0.11 | τ子 | 1.9 |
| 电子中微子 | $<10^{-9}$ | μ子中微子 | $<10^{-4}$ | τ子中微子 | $<10^{-3}$ |
| 上夸克 | 0.0047 | 粲夸克 | 1.6 | 顶夸克 | 189 |
| 下夸克 | 0.0074 | 奇异夸克 | 0.16 | 底夸克 | 5.2 |

注：实验已经确认中微子的质量不为零，但是还无法准确地测得其确切值。

在弦论中，真正的基本元素只有一种——各种不同种类的粒子不过是弦所能激发的不同振动模式。我们可以用常见的小提琴或大提琴的弦来加以说明。大提琴的弦有很多种振动模式，不同的振动模式对应着不同的音符。就是依靠这些不同的振动模式，大提琴才能演奏出各种不同的声音。弦论中的弦也是如此：这些弦也有着不同的振动模式，只不过这些振动模式对应的不是各种不同的声音；弦论中不同的振动模式对应着不同的粒子。需要认识到的关键之处在于，弦的某种特定振动模式产生的是某一特定的质量、特定的电荷、特定的自旋，等等——正是这些性质上的不同，使得一个粒子不同于另一个粒子。按某种模式振动的弦可能具有电子的性质，而按另一种不同模式振动的弦可能具有的是上夸克的性质，也可能是下夸克的性质，或者是表12.1中任何一种粒子的性质。构成电子的并不是“电子弦”，构成上夸克或者下夸克的也不是“上夸克弦”或者“下夸克弦”。唯一的一种弦就可以形成种类繁多的粒子，因为弦的振动模式种类繁多。

你或许明白了，弦论的这一特点意味着向统一迈出了一大步。如果弦论真的是正确的话，那么表12.1中那令人头晕目眩的粒子表所表示的就只是一种基本成分的不同振动模式。单独一种弦演奏出来的不同音符可以解释已观测到的所有粒子。在超微观尺度上，宇宙演奏了一曲弦交响乐来将所有的物质化为实在。

用弦论的方式解释表12.1中的粒子非常美妙。不过，弦论还能够让我们在统一之路上走得更远一些。在第9章以及前面的有关内容中，我们曾经讨论过大自然中的力在量子水平上是如何通过交换粒子来传递的，这些信使粒子可见表12.2。弦论中的信使粒子就像弦论中的物质粒子一样。也就是说，每一种信使粒子都是弦的某种振动模式。光子是弦的一种振动模式；W粒子是弦的另一种振动模式；胶子也是弦的一种特定的振动模式。还有，最重要的一点，施瓦茨和谢尔克在1974年发现的特别振动模式具有引力子的性质，因而引力也被包括到弦论的量子力学框架下了。这样一来，不仅物质粒子，还有信使粒子——甚至是引力的信使粒子——都来自于弦的振动。

**表 12. 2　自然界中的 4 种力，以及传递这 4 种力的粒子的名称和质量**

| 力 | 传递力的粒子 | 质　量 |
|---|---|---|
| 强 | 胶子 | 0 |
| 电磁 | 光子 | 0 |
| 弱 | W，Z | 86，97 |
| 引力 | 引力子 | 0 |

注：表中的数值是通过与质子质量比较所得。实际上有两种W粒子，所带电荷分别为＋1和-1，质量相同；为简化起见，我们略掉这一细节而只说存在W粒子。

综上所述，弦论不仅仅是第一个成功将引力和量子理论合并起来的理论，还是一个能够统一描述所有物质和所有力的理论体系。这就是为什么20世纪80年代中期上千名理论物理学家从他们的老本行中抽出身来，投入到弦论的研究中的原因。

为什么弦论会有用

弦论得以发展之前，科学进展的途中到处是合并引力与量子力学的失败之举。究竟是什么原因使得弦论能够获得这样巨大的成功呢？我们已经讲过施瓦茨和谢尔克是如何惊奇的认识到，按某种特别模式振动的弦具有引力子的性质，因而两人提出弦论是一个可以用来合并

引力和量子理论的现成框架。从历史发展的角度看，这就是弦论偶然的降临人世的过程；但是，为什么只有弦论能够成功而其他的尝试均以失败告终呢？这值得我们进一步思考。图12.2展示的就是广义相对论和量子力学的矛盾——在超小的距离（时间）尺度上，量子不确定性变得如此严重以至于广义相对论所依托的平滑几何模型不再成立——现在的问题是，弦论是怎么解决这一矛盾的？难道弦论能够平复超小尺度上时空的猛烈涨落吗？

弦论主要的新特征在于其基本成分不再是一个点粒子——没有尺寸的点——而是有空间延展性的客体。这一点正是弦论能够成功合并引力与量子力学的关键。

图12.2所示的猛烈涨落起源于将不确定原理应用到引力场；随着尺度越来越小，不确定原理使得引力场的涨落变得越来越大。在超小尺度上，我们用引力子来描述引力场，这就好像我们在分子的尺度上用$H_2O$分子描述水。在这种框架下，引力场的猛烈涨落可以看做是大量的引力子狂乱地飞来飞去，就像强大的龙卷风卷起泥土沙石一样。如果引力子是点粒子（弦论之前，所有试图合并引力与量子力学失败之举都是基于这一观念），图12.2实际反映的是这些引力子的集体效应：距离尺度越小，躁动就会越猛。弦论改变了这一结论。

在弦论的框架下，每一个引力子都是一个振动的弦——不是点，长度大约为普朗克长度（$10^{-33}$厘米）。[12]既然引力子是引力场最精细最基本的成分，那么谈论小于普朗克长度的引力场行为就毫无意义。你的电视机屏幕的分辨率受像素大小限制，弦论中引力场的分辨率也受引力子尺寸的限制。因而，弦论中引力子（其他的一切也是如此）的非零尺寸为引力场的分辨率设定了一个极限，这个极限大约是普朗克尺度。

认识到这一点非常重要。图12.2中那不可掌控的量子波动的起源是我们将量子不确定性应用到任意小的尺度上——比普朗克长度还小

的尺度上。在基于点粒子的理论中，这样使用不确定原理毫无问题；但是我们也看到了，这样的应用会把我们带到广义相对论失效的境地。但是基于弦的理论则有一个内置的保护措施。弦论中，弦就是最小的成分，所以我们的微观之旅到了普朗克长度——也就是弦的长度——也便到了尽头。在图12.2中，第二高的那层代表的就是普朗克尺度；我们可以看到，在这一尺度上，空间结构仍有波动，因为引力场还是要服从量子涨落。不过这里的涨落已经足够温和，不会与广义相对论产生不可挽回的冲突。广义相对论的数学部分必须适当修改以包括这些量子波动，这种修改不会带来数学上的麻烦。

总之，通过限制最小尺寸的“小”，弦论限制了引力场量子涨落的“大”——这个大刚好使得量子力学与广义相对论不会发生灾难性的冲突。就是这样，弦论调和了量子力学与广义相对论的矛盾，并且有史以来的一次，将两者合并起来。

小尺度上的宇宙结构

更为广义的空间和时空的超微观性质意味着什么？首先，关于时空的传统概念必然会受到挑战；在传统概念中，空间和时间的结构具有连续性——你总可以连续的切割两点之间的距离或者两个时刻间的时间间隔，你可以一次又一次地将它们一分为二，无穷无尽。现在，你必须放弃这样的连续性概念；你不停地切割时空，最后总会达到普朗克长度（弦的长度）和普朗克时间（光走过弦长所用掉的时间），这个时候你会发现你无法继续分割空间和时间。一旦你达到宇宙最小成分的尺度时，“变得更小”这个概念便失去了意义。以无大小的点粒子为基础的理论体系中并没有这样的限制；但是弦是有尺寸的，所以弦论中有这样的限制。如果弦论是正确的话，关于时空的那些普通概念，我们所有日常生活所依赖的那些概念，在比普朗克尺度——弦本身的尺度——还小的水平上就不再有效。

至于在小于普朗克尺度的地方应该有什么新的概念，人们还未形

成一致的看法。有一种可能性同前面讲过的内容——即弦论如何将量子力学与广义相对论合并起来——相一致；普朗克尺度上的空间结构类似于格点或网格，格线之间的空间超出了物理的范畴。就像走在一块普通布料上的超小蚂蚁，它只能在两条线之间蹦来蹦去。或许超小尺寸上的运动也是如此，只能从空间的一条“线”蹦到另一条。时间也是颗粒状的结构。单独的时刻彼此靠得很近，但却不是连绵不断的。按这种方式思考的话，更小的空间和时间间隔的概念会在普朗克尺度上突然走到尽头。这就好比你总是可以把钱分成更小的份，可是最后，你总要面对一分钱，这个时候你突然就无法把钱继续分成小份了。超微观时空如果是格点结构的话，就根本不会有小于普朗克长度的距离或者小于普朗克时间的时间间隔这样的东西了。

另一种可能是，在极端的小尺度上，空间和时间并不是突然失去了意义；而是渐变地转成其他更加基本的概念。之所以不能说“变得比普朗克长度还小”这样的话，并不是因为你遇到了最基本的格子；而是因为空间和时间这样的概念变成了别的东西，因而你说“变得更小”时就像问9这个数是不是快乐一样无意义。也就是说，我们在宏观尺度上熟悉的空间和时间逐渐变成了超微观尺度上我们不熟悉的某种概念。他们的很多性质——比如长度和间隔——都变得毫无意义了。这就好比你可以研究液态水的温度和黏性——描述液体宏观性质所使用的概念——但是当你在单个$H_2O$分子的尺度上研究时，温度和黏性这些概念就变得毫无意义了；因而，尽管你可以在日常生活的尺度上一次又一次的分割空间和时间，但是当你来到普朗克尺度的时候发生了某种变化，这种变化导致分割这样的事情毫无意义。

包括我在内的很多理论物理学家都强烈地感觉到沿着这条路走下去可能会得到一些成果。但是，只有找出空间和时间转变成了什么更加基本的概念，我们才能走得更远。[60]到目前为止，这仍是未解之谜。不过，在一些研究工作（我们会在最后一章中加以讨论）中已经提出了一些意义深远的可能性。

更小的点

讲到这里，看起来任何一位物理学家都很难抗拒弦论的诱惑。我们终于有了弦论这样一个理论，它不仅仅承诺要实现爱因斯坦的梦想；它还能调和量子力学与广义相对论之间的矛盾；它用振动的弦来描述世间万物从而将所有的物质和所有的力统一起来；在弦论的世界中，超微观尺度上的空间和时间像转轮拨号电话一样好玩；一言以蔽之，弦论是一个能将我们对于宇宙的理解提升到一个全新层次的理论。但千万别忘了，还没有人看到过弦，而且除了我们将要在下一章中讨论的一些稀奇想法，即使弦论是正确的，人们也很可能永远都看不到弦。弦实在是太小了，直接观测弦就像是从100光年以外阅读现在的这一页文字。直接测量弦对我们的技术提出了很高的要求，我们现有的分辨率再得提高百亿亿倍才有可能。一些科学家大声嚷嚷着弦论这样远超直接实验检验的理论只能算是哲学或神学领域的研究对象，它不是物理。

我要说这样的观点缺乏远见，或者说非常的不成熟。或许我们永远都不能直接测量弦，不过这没关系，科学史中到处都是只能用间接的方法检验的理论。[13]弦论并不谦虚，它的目标和许诺非常之大。这一点令人兴奋同时也有其意义，如果一个理论要成为关于宇宙的唯一理论，它就不能只在目前讨论的这种水平上马马虎虎的与现实世界匹配，它也应该在细微之处尽善尽美。正如我们马上就要讲到的，有一些办法可能可以检验弦论。

20世纪60～70年代的物理学家，在理解物质的量子结构和支配其行为的各种力（引力除外）方面迈出了非常大的一步。在实验结果与理论思考的双重推动下，人们得到了研究这些问题的理论框架，那就是粒子物理的标准模型。标准模型的基础是量子力学和表12.1中的物质粒子以及表12.2中传递力的粒子（标准模型并没有将引力纳入其中，因而需要忽略引力子。另外，标准模型中还有一种希格斯粒子没有在表中列出）。当然，这里的粒子都是点粒子。标准模型可以解释

世界上所有的原子对撞机上产生的数据，因而标准模型的作者得到了极高的荣誉。但是，标准模型有其局限性。我们已经讨论过弦论之前的各种理论并不能成功的调和引力与量子力学。除此之外，标准模型还有另外一些问题。

标准模型既不能解释为什么正好是表12.2中列出的那些粒子传递各种力，也不能解释为什么物质正好是由表12.1中列出的那些粒子组成。物质为什么有三代？每代为什么有那些粒子？为什么不是两代或一代？电子的电荷为什么是下夸克电荷的3倍？μ子质量为什么是上夸克质量的23.4倍？顶夸克的质量为什么是电子质量的350000倍？宇宙中为什么会出现这些看起来完全随机的数字？标准模型将表12.1和表12.2（忽略其中的引力子）中的粒子都当做输入参数，然后精确地预言粒子之间的相互作用和影响。就像你的计算器不能解释你所输入的数字，标准模型也不能解释它的输入参数——各种粒子及其性质。

思索这些粒子的性质并不仅仅是一个为什么种种神秘细节恰好是这样或那样的学术问题。过去百年间的科学实践使科学家们认识到，宇宙之所以具有人们日常经验所熟知的那些性质完全是因为表12.1和表12.2中的那些粒子恰好具有它们该有的性质。假如某些粒子的质量和电荷稍稍变化一点，使恒星发光放热的核反应过程可能就不会发生。没有恒星的宇宙完全是另外一个世界。因而，基本粒子的各种详细特性是与所有科学中最深刻的问题联系在一起的，这个最深刻的问题就是：基本粒子所具有的性质为什么恰好可以使核反应过程发生，恒星发光，行星得以围绕恒星而形成，而且其中至少有一颗行星上出现了生命？

标准模型完全回答不了这些问题，因为粒子性质只不过是标准模型的一部分输入参数。如果粒子性质不能确定下来，标准模型就无法运作，也给不出任何答案。而在弦论中，粒子的性质是由弦的振动决定的，因而弦论可以为粒子的种种性质提供一个解释。

弦论中的粒子性质

为了更好地理解弦论是如何解释各种问题的，我们最好先对弦的振动如何导致粒子的性质有一个更好的认识，所以我们先来看看粒子最简单的性质——质量。

从$E=mc^2$这个公式中，我们可以知道质量和能量可以彼此转化；这一点就像美元和欧元是可以互相兑换的一样（与货币的兑换略有不同的是，能量和质量按固定汇率兑换，这个汇率就是$c^2$）。我们的生活依靠的就是爱因斯坦方程。太阳每秒钟可以将430万吨的物质转化为能量，我们生活所需的光和热就是这些能量的一小部分。未来的某一天，我们或许可以仿效太阳的方式在地球上安全的利用爱因斯坦方程，到那一天，人类或许就可以获得无穷无尽的能量了。

在上面的这些例子中，能量来自于质量。爱因斯坦方程也可以反过来用——也就是说将能量转化为物质——而这正是弦论使用爱因斯坦方程的方式。弦论中，粒子的质量不是别的，正是弦的振动能量。例如，弦论是这样解释一个粒子为什么会重于另一个粒子的：构成较重粒子的弦比构成较轻粒子的弦振动得更加快速也更加猛烈。更快更猛的振动意味着更高的能量；而根据爱因斯坦方程，更高的能量意味着更大的质量。反过来说，一个粒子的质量越轻，也就意味着弦振动得越慢越平和；而无质量的光子和引力子则对应着弦可能有的最平静温和的振动模式。61[14]

粒子的其他性质，比如电荷和自旋等与弦的振动的其他一些更加深奥的性质有关。与质量相比，这些性质很难不用数学就加以描述，但基本思想是一样的：振动模式就是粒子的指纹；我们用来区分粒子的所有性质都由弦的振动模式决定。

20世纪70年代早期，物理学家曾分析过弦论的最初化身——玻色型弦论——的振动模式以便确定理论中预言的粒子的性质，但是他们

遇到了一些麻烦。玻色型弦论中的每种振动模式都具有整数自旋：自旋0，自旋1，自旋2，等等。这是一个很大的问题，虽然传递力的粒子的自旋正是整数；但是物质的粒子（比如电子和夸克）的自旋并不是整数，这些粒子的自旋为半整数，即1/2。1971年，佛罗里达大学的皮埃尔·雷蒙德决定攻克这一问题。雷蒙德很快就找到了一种修改玻色型弦论方程的办法，修改后的方程可以将半整数的振动模式纳入其中。

事实上，仔细地考查雷蒙德的研究，以及施瓦茨和他的合作者安德烈·内沃发现的结果，还有稍后一些的费迪南多·格里奥奇、乔·谢尔克和大卫·奥利弗的发现，人们认识到修改后的弦论中不同自旋的振动模式之间存在一种完美平衡——一种新颖的对称性。研究者们发现，新的振动模式按自旋相差1/2的方式成对出现。每一种自旋1/2的振动模式有自旋0的振动模式伴随；每一种自旋1的振动模式有自旋1/2的振动模式伴随。整数自旋与半整数自旋之间的对称性称为超对称性，于是，超对称弦论（简称超弦）诞生了。10年之后，施瓦茨和格林就是在超弦的框架下证明了所有威胁弦论的可能反常最终相消。因而，施瓦茨和格林的论文所引发的弦论革命更适合被称为第一次超弦革命（在后面的内容中，我们常常会提到弦和弦论，当我们这么说的时候，我们实际上指的是超弦和超弦理论）。

有了这些基础，我们可以脱离泛泛的讨论，仔细看看弦论关于这个宇宙究竟说了些什么。事情很清楚：弦所激发的各种振动模式中，必然有一些振动模式的性质与已知的粒子相符合。理论中有自旋1/2的振动模式，弦论必须使自旋1/2的振动模式与表12.1中所列出的已知物质粒子精确符合。理论中也有自旋1的振动模式，弦论也必须使自旋1的振动模式与表12.2中所列出的已知信使粒子精确符合。最后，如果实验上真的发现了自旋为0的粒子，比如希格斯场所预言的粒子，那么弦论就必须使自旋为0的振动模式与实验上发现的那些粒子的性质精确符合。总之，弦论要想成为一个正确的理论，它的振动

模式必须能够解释标准模型的粒子。

弦论的机会来了。如果弦论是正确的话，它就能够解释实验上测得的粒子的性质，各种性质不过是弦所能激发的振动模式。如果弦的振动模式能够与表12.1和表12.2所列的粒子性质相匹配的话，那么不论实验上是否能够直接观测到弦的存在，我相信即使那些弦论最苛责的刁难者也会开始相信弦论。如果理论真的能与实验符合得很好的话，那么弦论不但能成为人们长久以来一直追寻的统一理论，还能够首次给予宇宙为什么是现在这个样子这一问题一个真正的基本解释。

那么弦论是如何应对这一挑战的呢？

振动太多了

弦论的第一次登场以失败告终。对于最早的那批先驱者来说，弦的不同振动模式有无数种，图12.4所示的就是这一无穷级数的最初几项。但在表12.1和表12.2中，只有有限的几种粒子存在，因而从一开始我们就很难和现实世界匹配得上。更为严重的是，如果我们利用数学来分析这些振动模式可能的能量——也就是质量——的话，我们会发现理论和实验观测存在着另一个明显的分歧。弦可能的振动模式的质量与表12.1和表12.2中所列出的实验观测值并不一致。我们现在来看看这是为什么。

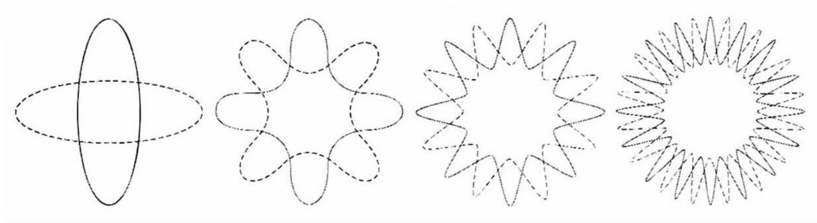


图12.4　最初的几种振动模式。

早在弦论刚刚诞生的日子，物理学家们就了解到弦的硬度反比于

弦的长度（更准确地说是长度的平方）：越长的弦越容易弯曲，越短的弦会变得越硬。1974年，施瓦茨和谢尔克提出降低弦的尺寸以使之实现引力的强度；这种做法同时也导致了弦的张力增强，简单的分析表明弦中的张力大约是千万亿亿亿亿（$10^{39}$）吨，这个数值是钢琴中的弦的张力的100000000000000000000000000000000000000000（$10^{41}$）倍。想象一下吧，如果你想把一根细小又极其硬的弦弯成图12.4中的那些形状，你会发现峰和槽的数目越多的话，你所要花费的力气也就越大。反过来看，如果一根弦以这样一些形状振动，它所能释放出来的能量也将异常巨大。因而，除了最简单的振动模式，弦其他所有的振动模式都意味着超高能量，根据$E=mc^2$，这也就意味着那些振动模式的质量超级巨大。

当我说巨大这个词的时候，我想表达的是真正的巨大。计算表明，弦振动的质量是一个级数，所有的质量都是一个基本质量的倍数，这个基本质量就是普朗克质量；这一点就像音乐中的和弦一样，泛音的频率都是基频的倍数。按粒子物理的标准，普朗克质量实在太过巨大了——差不多是质子质量的千亿亿倍（$10^{19}$），几乎是尘埃颗粒或者细菌那么重了。而弦振动的质量只能是普朗克质量的0倍、1倍、2倍、3倍等，这一事实表明，除了弦的0质量振动模式，所有的振动模式的质量都太过巨大了。[15]

如你所见，表12.1和表12.2中的个别粒子无质量，但是大部分粒子具有质量。相比于普朗克质量，这些有质量的粒子其质量非常之小，简直比文莱的苏丹需要借贷的概率还小。因而，我们可以很明显地看出已知粒子的质量并不满足弦论所预言的质量。这是否意味着弦论失败了呢？你或许会这么想，不过事实并非如此。无穷种振动模式的质量远大于已知粒子的质量，这一点的确是弦论必须战胜的一个挑战。多年的研究表明，有一些办法可能会帮助弦论跨越这一难关。

首先，实验告诉我们越重的已知粒子就越不稳定；重的粒子常常

很快衰变成低质量粒子雨，这些低质量粒子再衰变，最终留给我们的是表12.1和表12.2中最轻、最为人们熟悉的粒子（举个例子，顶夸克会在$10^{-24}$秒内就衰变掉）。我们期望这样的机制对“超重”的弦的振动模式也成立，那样的话，我们就能够解释为什么在高热的早期宇宙中即使有大量的超重振动模式产生，在今天的宇宙中我们也看不到它们了。即便弦论是正确的，我们也很难看到弦的超重振动模式；或许我们唯一的机会是粒子加速器上的高能碰撞。但是，目前的加速器能量大概只是质子质量的1000倍；相比于弦论中的非最小振动模式（最小的振动模式对应着零质量的粒子），这样的能量实在是太微弱了。因而，弦论所预言的粒子塔中每一个质量都算得上很大，甚至最小的质量都要比目前技术水平能达到的能量大千万亿倍。所以弦论与实验观测并不矛盾。

从这一解释中我们可以很明白的看清一个事实：弦论与粒子物理唯一的接触机会就是弦的低能振动模式——也就是无质量粒子，因为其他的振动模式的质量都远远超越了当前技术的掌控范围。既然这样，我们就要问：为什么表12.1和表12.2中的大部分粒子都是有质量的呢？这是一个非常重要的问题，不过它并没有乍看之下那么可怕。普朗克质量非常巨大，即使目前已知的最重粒子——顶夸克，其质量也仅仅是普朗克质量的0.0000000000000000116（大约$10^{-17}$）倍；至于电子，它的质量仅仅为普朗克质量的0.0000000000000000000000034倍（约为$10^{-23}$）倍。所以，一阶近似下——直到$10^{17}$分之一都是有效的——同普朗克质量相比，表12.1和表12.2中的所有粒子都是零质量（一阶近似下，同文莱的苏丹相比，地球上的大多数人的财富都是零），正如弦论所预言的那样。我们的目标是改善这一近似，并且用弦论解释表12.1和表12.2中的粒子质量为什么会和零有一个小小的偏离。由此可见，无质量的振动模式与实验上的数据并不一致这一事实，并不像你一开始想象的那么严重。

虽然情况已经令人大受鼓舞，但是精细的分析并非易事。利用超弦理论的方程，物理学家们可以写出所有的无质量振动模式。其中之一就是自旋为2的引力子，而我们知道正是它的成功唤起了以后的研究；引力子的存在使得引力可以是量子弦论的一部分。计算告诉我们的另一件事是：自旋为1的无质量振动模式比表12.2中列出的粒子多很多；自旋为1/2的无质量振动模式也比表12.1中列出的粒子多很多。更为糟糕的是，自旋为1/2的无质量振动模式并没有表现出任何具有代的结构的迹象。更加仔细地分析后，人们发现，将弦的振动模式与已知粒子对应起来的确并非易事。

20世纪80年代中期时的情况就是这样，一方面，人们有理由因为超弦而激动万分；另一方面，人们又有理由对超弦保持怀疑。只有一点毫无疑问：超弦理论的确向着统一迈出了大胆的一步。超弦理论毕竟是第一个将引力和量子力学合并起来的自洽方法。它对物理学所做的一切就像罗杰·班尼斯特[62]与4分钟1英里（1英里≈1.6千米）：将不可能变为可能。超弦理论使我们相信，攻克20世纪物理学两大支柱之间的壁垒并非不可能。

当然，更进一步的分析，试图用超弦理论解释物质和自然界中的力的详细特质时，物理学家们遇到了困难。这些困难使得一些人质疑超弦理论；在怀疑者们看来，如果不提统一方面的潜在可能，超弦理论只不过是一种与物理的宇宙毫无关系的数学结构罢了。

在超弦理论的怀疑者们看来，前面讲过的问题并不是最紧要的，超弦理论的特性中有另外一个更为严重的弱点。现在我来介绍一下这个问题。超弦理论的确能够成功地将引力与量子力学组合起来，没有受困于数学上的不自洽，而数学上的不自洽却是之前的很多尝试失败的原因。但是，在超弦理论最初的几年里，物理学家们发现，如果宇宙有3个空间维度，超弦理论的方程就无法拥有那些令人羡慕的性质；仅当宇宙有9个空间维度的时候，超弦理论才具有数学上的自洽性。这也就是说，加上时间维度的话，超弦理论中的宇宙必须拥有十

维时空。

与这个听起来非常古怪的要求相比，将超弦的振动模式与已知粒子种类精确的对应起来这个难题只能算是一个二等问题。超弦理论要求存在另外6个没人见过的空间维度。这可不是什么好事，这是一个问题。

这真的是一个问题吗？

早在超弦诞生之前，20世纪最初几十年的理论发现表明，额外维度根本不必成为一个障碍。而且，在20世纪末的更新版本中，物理学家们证明了额外维度有能力在弦论的振动模式与实验上发现的基本粒子之间搭起一座桥梁。

下面我们就一起来看看这赏心悦目的理论进展。

在更高的维度中统一

1919年，爱因斯坦收到了一篇论文。这篇很容易因为被当做奇思怪想而遭遗弃的论文出自没什么名气的德国数学家西奥多·卡鲁扎之手。在短短的几页纸上，作者提出了一种统一当时已知的两种力——引力和电磁力——的方法。为了达到统一的目的，卡鲁扎抛出了一种相当激进的方案；这一方案明显违背了一些非常基本、可以完全想当然，甚至不需要质疑的东西。卡鲁扎提出，宇宙并不是只有3个维度。他恳请爱因斯坦和物理学家们接受宇宙有4个空间维度的可能性。这样一来的话，连同时间一共就有了5个时空维度。

我们立即要问，这样做究竟意味着什么？我们说有3个空间维度时，我们指的是存在着3个独立的方向或轴，而你可以沿着这些方向运动。在你当前的位置，你可以定出左右、前后，上下这些方位。在一个三维的宇宙中，你所做的任何运动都是沿着这3个方向分别运动的某种组合。我们也可以这样说，在一个三维宇宙中，你需要3条信

息才能确定一个位置。比如，在一座城市中，你需要知道某栋建筑的街道以及跟它相交的街道，还有具体的楼层才能搞清一场晚宴到底在哪里举行。如果你希望客人们按时到达，那么你还需要第4条信息：时间。这就是我们说时空具有4个维度的原因。

而卡鲁扎提出，除了左右、前后、上下，宇宙还有另外一个空间维度；由于某种原因，人们无法看到这个额外的维度。如果真是这样的话，那就意味着事物还有另外一个独立的方向可以运动。因而我们需要4条信息才能在空间中定位。如果算上时间的话，我们需要的就是5条信息。

就是这样，爱因斯坦在1919年4月收到的那篇论文说的就是这个事情。问题在于，爱因斯坦为什么没把这篇论文扔掉呢？既然我们没有看到另外的维度——我们从来没有因为一条街道、一条交叉街道和楼层号这3条信息还不足以帮助我们找到想去的地方而懵懵懂懂——我们为什么要在乎这个古怪的想法？现在我们就来看看为什么。卡鲁扎发现，爱因斯坦的广义相对论方程在数学上可以相当容易的推广到具有更高空间维度的宇宙中。卡鲁扎进行了这样的扩充；然后很自然地发现，扩充后的广义相对论的高维版本并不仅仅包括了原始的广义相对论方程，还因为更多的维度而有了一些新的方程。仔细研究这些多出来的方程后，卡鲁扎发现了一些非常奇特的东西：这些方程居然是19世纪麦克斯韦发现的用以描述电磁场的方程！为宇宙添加了一个新的维度后，卡鲁扎竟然解决了爱因斯坦眼中所有物理学中最重要的一个问题。卡鲁扎提出的理论体系可以将爱因斯坦的原始广义相对论方程和麦克斯韦的电磁场方程组合起来。这就是爱因斯坦没有扔掉卡鲁扎论文的原因。

直观上，你可以这样理解卡鲁扎的理论。在广义相对论中，爱因斯坦重新认识了空间和时间；空间和时间的扭曲和拉伸，可以使引力以一种几何式的方式现身。卡鲁扎在他的论文中提出，空间和时间的几何疆界可以延伸得更远。爱因斯坦认识到引力可以看做是普通三维

空间和一维时间中的蜷曲和涟漪；而卡鲁扎发现，再加上一维空间的宇宙中会有更多的蜷曲和涟漪，而这些可以使他获得描述电磁场的方程。在卡鲁扎的手中，爱因斯坦用几何描述宇宙的方法强大到足以统一引力和电磁场。

当然，问题还是存在的。尽管数学上没有问题，但是却没有——现在也没有——任何证据证明存在着比我们熟知的三维还多的空间维度。那么卡鲁扎的发现只是理论上的一个意外呢？还是的确与我们的宇宙有着某种不为人所知的联系呢？卡鲁扎本人对自己的理论深信不疑——比如，他曾经研究过一篇游泳方面的文献，以便学会游泳潜入深海中——但是一个看不见的空间维度这样的想法，不管它在理论上是多么的引人注意，实际上总是令人难以接受。1926年，瑞士物理学家奥斯卡·克莱因为卡鲁扎的想法注入了新的活力，他发现了一种隐藏额外维度的办法。

隐藏的维度

为了理解克莱因的想法。我们先来想象一下这样一个场景：菲利普·帕迪特[63]在珠穆朗玛峰与洛子峰[64]之间走钢索。我们从很多千米以外看到的这个场景，就像图12.5所示意的那样。钢索看起来就像条一维绳子——只在其长度的方向上延展。这时有个人告诉我们一条蚯蚓正在菲利普的前面一点慢慢地爬，我们这时只能为蚯蚓祈祷了，希望这个可怜的小家伙能够一直在菲利普的前面以免遭灭顶之灾。稍稍回过点神后，我们都认识到绳子并不是只有我们能看见的左右这个维度。尽管只用肉眼我们很难从远处看清楚，但是除了长度的方向，绳子的确还有另外一个维度：那就是缠绕在绳子上的“蜷曲”维度。我们在望远镜的帮助下看清了弯曲的维度，我们看到蚯蚓并不是只能在“长”的方向上左右爬动，它也能在很“短”的方向上，绕着绳子顺时针或逆时针爬动。也就是说，在绳上的每个点，蚯蚓都有两种独立的方向可以爬动（而这也正是我们说绳子的表面是二维时想表达的意思[65]）。因而，要想避开菲利普的脚，蚯蚓有两种选择：要么像我们之

前认为的那样，始终爬在菲利普的前面；要么绕着绳子爬到下边，让菲利普先过去。

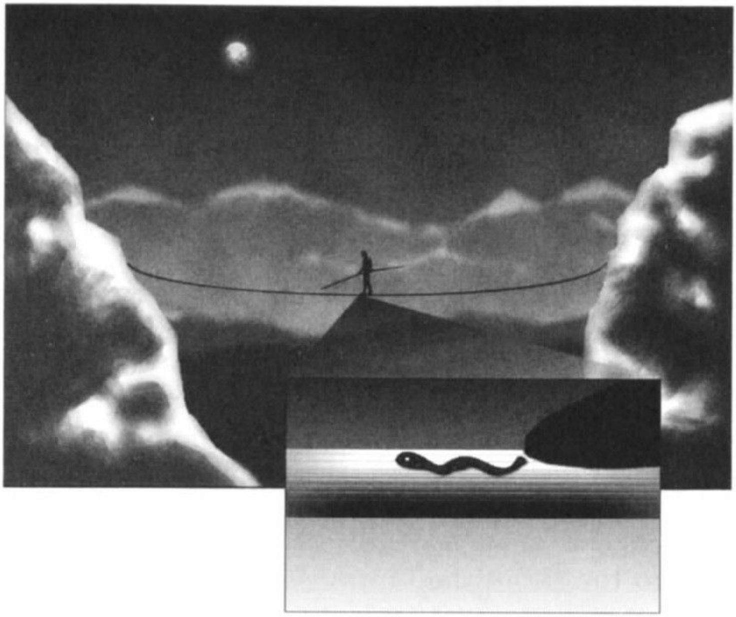


图12.5　从远处看，拉紧的绳索好像是一维的。只有用很强的望远镜才能看到它蜷曲的第二维。

绳子向我们展示了维度——物体可以在其上移动的独立方向——可以有两种定性上完全不同的种类。维度既可以大到我们能够看见，就像前面所说的沿着那根绳索表面的左右维度；也可以小到我们很难看见，就像环绕绳索表面的顺时针逆时针的那个维度。在我们所举的这个例子中，看到绳索表面小小的环绕维度并不难。我们所需的只是一台有效的放大设备。我们也很容易理解，蜷曲维度越小的话，看到它们也就越难。从几千米远的地方，看到一条钢索的蜷曲维度是一回事；要想看到一根牙线或者神经纤维的蜷曲维度则是另外一回事。

克莱因的贡献之处是：他首先提出，对于宇宙中的某个物体正确的事很可能对于宇宙本身也是正确的。也就是说，绳索的表面既可以

有很长的维度也可以有小到很难看见的蜷曲维度，宇宙也是如此。或许我们熟知的3个空间维度——左右、前后和上下——就像绳索的水平维度，属于很容易看见的大维度。另一方面，就像绳索有很小且又蜷曲的环形维度，空间的结构可能也有很小且还蜷曲的环形维度，这一维度可能如此之小以至于我们现有的各种放大设备还不足以发现它们的存在。克莱因提出，由于尺度非常之小，这些维度成了隐藏的维度。

那么这些很小的维度究竟可以有多小？如果把量子力学引入克莱因的原始理论中的话，数学分析可以告诉我们这一额外的环形空间维度的半径可能只有普朗克长度那么长，[16]这么小的长度显然是实验达不到的水平（当今世界最高水准的实验设备只能探测到千分之一原子核大小的长度，而普朗克长度比这个长度还小千万亿倍）。但是对于普朗克尺度的蚯蚓来说，这个又小又蜷曲的环形维度足够它在上面溜达了，正如图12.5中所示的那种绳索足够一条真正的蚯蚓绕着它爬来爬去。当然，真正的蚯蚓会发现在蜷曲的环形维度上转来转去没什么意思，因为没走多远就回到了起点；而这一点对于普朗克尺度的蚯蚓来说也很烦恼。不过，除了不够长这一点外，这一蜷曲的小小环形维度看起来跟普通的三维平直维度没什么分别。

为了对这个问题有一个直观的印象，我们必须注意到我们所说的蜷曲维度——顺时针逆时针这个方向——在沿着绳索延展维度的每一点都存在。蚯蚓可以在沿着绳索延展方向的每一点绕着环形维度爬动。所以，我们可以说绳索表面可以描述为有一个长长的维度，且在这个长长的维度的每一点上都有一个小小的环形维度，如图12.6所示。把这一点记在脑中，因为克莱因就是用这样的办法隐藏了卡鲁扎的额外空间维度。

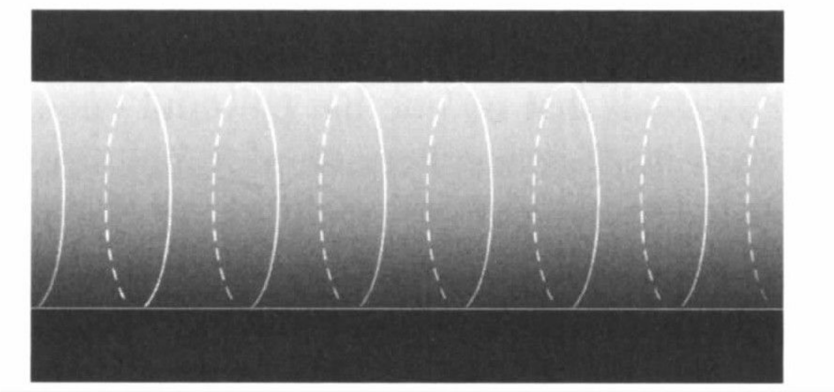


图12.6　拉紧的绳索表面，长长的一维的每一个点上都绕着一个维度。

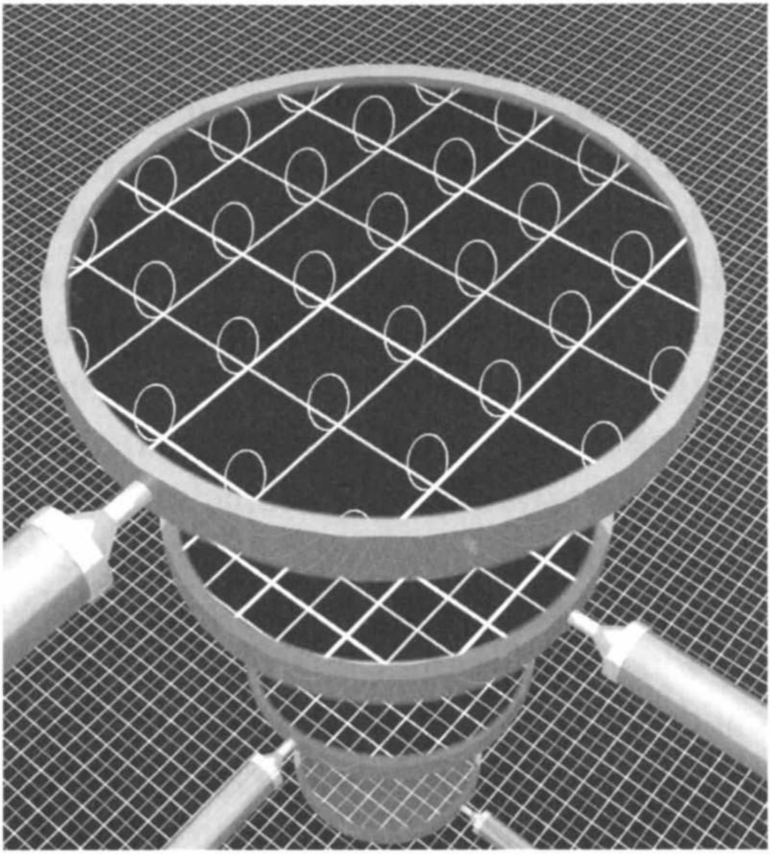


图12.7　卡鲁扎克莱因方案说的是，在非常小的尺度上，每一个点都附着一个额外的蜷曲维度。

为了搞清这一点，我们再来一点一点的查看越来越小尺度上的空间结构，如图12.7所示。首先，在放大倍数不多的地方，什么新东西也没有：空间仍然只是普通的三维结构（我们在书中用两维的格子示意性的代表）。但是，当我们到达普朗克长度——就是图中放大的最高层——的时候，克莱因提出，一种新的蜷曲的维度出现了。正如沿着绳索延展方向的每一个点上都存在着一个环形维度；克莱因提出的环形维度也存在于我们熟知的三维空间的每一个点上。图12.7中，我们示意性的在沿着延展方向的每一个点上画了一个额外的环形维度

（我们只能用格子示意性的说明，要是真的在每一个点都画一个圆环，那我们在这张图上就什么也看不清了），你可以看出这与图12.6中的绳索多么的类似。因此，在克莱因的方案中，空间有3个普通的平直维度，在这3个维度的每一个点上都有一个额外的环形维度。注意，额外维度并不是普通三维空间中的一个圈；或许这张示意图的局限性会使你这么想，但是千万别，因为不是这样。额外维度是一个全新的方向，与我们所熟知的3个方向完全不同，它存在于普通三维空间的每一个点上，因为太小所以逃过了我们迄今为止最高级的设备的探测。

这样修改了卡鲁扎的原始思想之后，克莱因回答了宇宙是如何获得三维之外的维度并将它们隐藏起来这个问题。从此之后，这一理论正式被称为卡鲁扎克莱因理论。我们还记得，只要有一个额外的空间维度，卡鲁扎就可以将广义相对论与电磁场融合起来；因而，卡鲁扎克莱因理论看起来明显就是爱因斯坦想要的理论。事实上，爱因斯坦和其他一些人的确因为通过一个全新的隐藏维度实现了统一而激动万分，人们花了大量力气来仔细研究这一理论。不久，人们就发现卡鲁扎克莱因理论有自己的麻烦。最明显的一点是，将电子纳入额外维框架下的努力均以失败告终。[17]爱因斯坦本人直到20世纪40年代早期还偶尔涉猎卡鲁扎克莱因体系，但是这一理论由于最初的许诺迟迟不能兑现而逐渐淡出了人们的视野。

不过，几十年之后，卡鲁扎克莱因理论气势汹汹地卷土重来了。

弦论与隐藏的维度

除了试图解释微观世界时遭遇难题，还有另外一个原因使得科学家们在面对卡鲁扎克莱因理论时犹豫不前。很多人觉得卡鲁扎克莱因理论在假定隐藏的空间维度时太过随意，有太多任意性。克莱因并不是通过严格的推导一步步得出新的空间维度这一想法。相反，他就像变戏法一样突然从帽子中拿出了他的想法，之后的分析使他不经意间

发现了广义相对论与电磁场的联系。因而，尽管这个发现本身非常伟大，但却不带有某种必然性。如果你问卡鲁扎和克莱因为什么宇宙非得有5个维度，而不是4个、6个、7个，或者是7000个，他们大概只能搪塞你一句：“为什么不是5个呢？”

但是30年之后，一切极速扭转。弦论成了第一个成功融合广义相对论和量子力学的方法；而且，弦论甚至有机会统一所有的物质和所有的力。但是，弦论的方程不但在四维时空中不起作用，在五维、六维、七维，甚至七千维中都不起作用。由于一些我们将要在下节讨论的原因，弦论的方程只在十维时空——九维空间加一维时间——中才起作用。弦论要求有更多的维度。

这是一个本质上全然不同的结果，人们从未在之前的物理学史中遇到这一情况。弦论之前，从没有任何理论说过宇宙应该有多少空间维度。从牛顿到麦克斯韦到爱因斯坦的每一个理论都假定宇宙有3个空间维度，就像我们假定太阳明天还会升起那么自然。卡鲁扎和克莱因通过提出有4个空间维度的方式首次挑战了这一假定，但却带来了另一个假定——4个空间维度这个假定，虽然不同于3个空间维度的假定，但毕竟还是一个假定。而现在，弦论的方程首次预言了空间维度的数目。在弦论中，是计算——而不是假定、假设，甚至充满灵感的猜测——决定了空间维度的数目；但令人惊讶的一点是，算出来的空间维数竟然不是3，而是9。弦论不可避免的将我们带到了6个额外空间维度的宇宙，从而为唤醒卡鲁扎克莱因理论提供了现成的背景。

原始的卡鲁扎克莱因理论只假定存在一个额外维度，不过很容易就能推广到2个、3个，甚至弦论所要求的6个额外维度。比如，我们可以将图12.7中额外的一维环形换成图12.8（a）中的二维球面形（回想一下第8章中的有关讨论：球面是二维的，因为你只需要两条信息——比如地球表面的纬度和经度——就可以确定位置）。与讨论圆环时一样，你需要将球面想象成附着在普通三维空间的每一点上。为了使图12.8（a）易于辨认，我们只是示意性的用网格来表示三维

空间。在这样的宇宙中，你需要5条信息才能定位：3条信息用来确定你在大维度中所在点的位置（比如所在街道、交叉街道和楼层这3条信息），另外两条用来确定你在该点的球面上所在的位置（需要知道经度和纬度）。当然，如果额外维度的半径足够小——比原子半径小数十亿倍——那么在考虑相对很大的事物，如我们自己时，最后两条信息就无关紧要。不过，要是考虑的是超微观尺度上的事物，我们就必须将5个维度一起考虑。我们需要全部的5条信息才能定位超微观尺度蚯蚓。如果再算上时间的话，总共就需要6条信息才能在正确的时间赶到正确的地点参加晚宴。

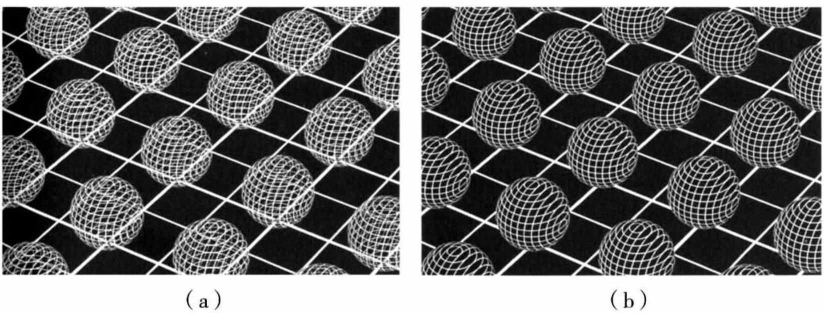


图12.8　从近处看一个普通三维的宇宙，这个用格子表示的宇宙具有额外的以空心球形式存在的两个蜷曲维度[图（a）]；或者以实心球形式存在的3个蜷曲维度[图（b）]。

我们再来深入一下。在图12.8（a）中，我们只考虑了球形的表面。现在我们来想象一下图12.8（b）中的情形，这时的空间结构也包括球的内部区域——普朗克蚯蚓这回可以钻到球里面去了，就像普通的蚯蚓可以钻到苹果里面一样。于是我们需要6条信息才能明确蚯蚓的位置：3条信息确定它在普通的三维空间中所在点的位置，另外3条确定它在该点的球中的位置（经度、维度、掘进深度）。再加上时间，这就是一个七维宇宙的例子。

现在我们需要跳跃一下了。虽然很难画出，但是我们可以想象一下。我们需要想象在普通的三维空间的每一个点，并不仅仅像图12.7

那样有一个额外维，或是像图12.8（a）那样有两个额外维，又或是像图12.8（b）那样有3个额外维，而是有6个额外的空间维度。我承认我画不出这样的图，我所见过的人里也没有人能画出这样的图。但是它的意思很清楚。要想在这样的一个宇宙中确定普朗克蚯蚓的位置，我们需要9条信息：3条用以确定它在普通三维空间中所在点的位置；另外6条确定它在该点的六维蜷曲空间中的位置。再考虑到时间的话，这就是一个十维时空的宇宙，而这也正是弦论方程所要求的宇宙。如果额外的6个维度足够小的话，它们就能逃脱实验上的探测。

隐藏维度的形状

实际上，弦论的方程不仅仅决定了空间的维度，还可以决定额外维可能具有的形状。[18]在前面的几幅图中，我们只讨论了最简单的几种形状——环形、空心球面、实心球——而弦论方程选中的则是一类极其复杂的六维形，所谓的卡拉比丘形或卡拉比丘空间。这类形状是根据两位数学家尤金尼奥·卡拉比和丘成桐的名字命名的，这两位数学家早在弦论出现之前就在数学上发现了这类形状。图12.9（a）只是此类形状的一个粗略演示，要知道这张图是以二维演示六维，因此难免会有些失真。不过，我们还是能通过这张图对卡拉比丘形状有一个大概的认识。如果弦论中额外的6个维度真的是图12.9（a）中的卡拉比丘形，那么超小尺度上的空间就如图12.9（b）所示。在普通三维空间的每一个点上都有一个卡拉比丘形。你，我，还有所有的人都被这些很小的形状环绕甚至占据着。可以这样说，你从一处走到另一处的过程中，你的身体在所有的9个维度中穿行，一遍一遍的穿出进入这些形状。但是平均下来看，似乎你没有进入任何的额外维度中。

如果这些想法真的正确，宇宙的超微观结构就有着丰富的花样。

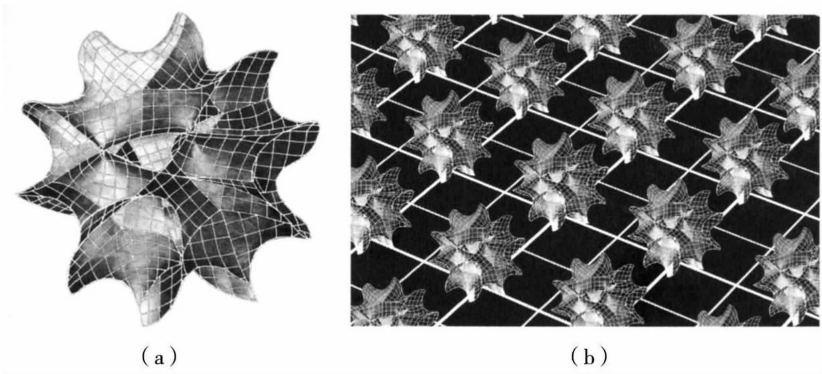



图12.9　（a）：卡拉比丘形的一个例子。（b）：具有卡拉比丘形额外维度的空间放大图。

弦物理与额外维度

广义相对论的优美之处在于引力的物理受空间几何的控制。考虑到弦论所提出的额外维度之后，你自然会想到几何控制物理的能力将大大增强。事实的确如此。我们首先来看一个我一直回避的问题：弦论为什么要求时空是十维？不用数学的话很难回答这个问题，不过还是让我尽量解释一下以便了解几何与物理是如何相互影响的。

想象一支被限制在平坦桌面的二维表面上振动的弦。这支弦可以产生很多不同的振动模式，但不管哪种模式，都是在桌子表面上向着左右或者前后振动。如果我们再允许这支弦在第3个维度中振动，也就是离开桌子表面向着上下振动，那无疑会产生另外一些振动模式。虽然我们很难画出在大于三维的空间中振动是什么样子，但这并不影响上述结论——更多的维度意味着更多的振动模式——的普适性。如果弦也可以在第4个空间维度中振动，那它就能产生出比只在三维空间中时更多的振动模式；如果弦还可以在第5个空间维度中振动，那它将能产生出比只在四维空间中时还多的振动模式，依此类推。认识到这一点非常重要，因为在弦论中有一个方程要求独立振动模式数目需要满足某种精确限制。如果这种限制被破坏，弦论的数学就会破产，它的那些方程将变得毫无意义。在一个只有3个空间维度的宇宙

中，振动模式的数目太小，因而无法满足那一限制；而在有4个空间维度的宇宙中，振动模式的数目还是太小；在有5个、6个、7个、8个空间维度的宇宙中，振动模式的数目都太小，但是在有9个空间维度的宇宙中，对弦的振动模式数目的限制被很完美的满足。正是因为这个原因，弦论需要有9个空间维度。[66][19]

这已经很好地展示了几何与物理的交汇作用，而将其与弦论联系起来则使我们得到的更多，事实上，为先前遇到的一个严重问题找到了一种处理办法。还记得吗？当物理学家们试图将弦的振动模式同已知粒子种类联系起来的时候，他们遇到了大麻烦。物理学家们发现，存在着太多无质量的弦振动模式，而且更为糟糕的是，振动模式的具体性质无法与已知的物质与力的粒子性质相匹配。但我在前面没有提到的是——因为那时我们还没有讨论到额外维——尽管这些计算考虑了额外维度的数目（部分的解释了为什么会发现这么多弦的振动模式），但却没有将额外维度的小尺寸以及复杂形状一道考虑，而只是假定所有的空间维度都是平直且是完全展开的，这就是一个重要的区别。

弦如此之小，以至于即使额外维蜷成卡拉比丘形，弦还是会在这些方向上振动。而这一点非常重要，原因有两个。首先，这样才能保证弦总是会在所有的9个空间维度中振动，从而使得对振动模式数目的限制始终都能够被满足，即使额外维度收紧到一起也没有关系。其次，正如吹入大号[67]中的气流的振动模式会受乐器本身形状的弯曲扭折影响，弦的振动模式也会受额外6个维度的几何形状的弯曲扭折影响。如果你将大号的形状变一变，比如将号管弄窄些或者将腔膛弄长些，气流的振动模式就会变化，从而使得乐器发出的声音有所变化。类似地，如果额外维度的形状及大小有所改变，弦的每一种可能的振动模式也都会受到很大的影响。而弦的振动模式决定着相应粒子的质量和所带的电荷，因此，额外维度在确定粒子性质方面扮演着重要角色。

认识到这一点非常关键。额外维度准确的大小和形状对弦的振动模式有着重大的影响，从而对粒子的性质也有重大影响。既然宇宙的基本结构——从星系和恒星的形成到我们所知的生命存在——敏感的依赖着粒子性质，我们可以说，宇宙的密码写在卡拉比丘形的几何之中。

我们在图12.9中看到了卡拉比丘形的一个例子，但要知道还有成百上千种卡拉比丘形。于是问题就成了到底是哪一种卡拉比丘形构成了时空结构中的额外维部分。这是弦论所需要面对的最重要的问题之一，因为只有明确了卡拉比丘形的具体形状，才能知道弦的振动模式的具体性质。可直到今日，这一问题仍未解决。原因在于，目前对弦论方程的理解还不能告诉我们怎样从很多卡拉比丘形中挑出一种；从已知方程的角度看，每一种卡拉比丘形都同样有效。而且这些方程甚至不能确定额外维度的大小。因为我们看不到额外维度，所以它们必须很小，但到底有多小则仍是我们所不知道的。

这算是弦论的致命缺陷吗？可能是吧，但是我并不这么看。我们在下一章中将会全面的讨论到，弦论学家们多年来一直无法抓到精确的弦论方程，他们所做的大量工作依靠的都是近似方程。这些近似方程已经告诉了我们很多弦论的特性，但是在某些问题上——包括额外维度的准确大小与形状上——近似方程的缺点展现无遗。随着我们进一步深化数学分析改进这些近似方程，确定额外维度的形式将成为我们的一个主要的——且可以达到的——目标。可惜目前为止，我们尚未能实现这一目标。

不过，我们还是可以问问是否有哪种卡拉比丘形可以使我们得到与已知粒子近似的弦的振动模式。这个问题的答案还是能令人感到欣慰的。

虽然我们远不能探索每一种卡拉比丘形，人们还是找到了能够带来与表12.1和表12.2大体相符的振动模式的某些卡拉比丘形。比如

说，20世纪80年代中期，菲利普·坎德拉斯、加里·霍洛维茨、安德鲁·斯特劳明格以及爱德华·威滕（这些科学家认识到卡拉比丘形对弦论的意义）发现卡拉比丘形中所包含的每一个洞——在精确定义的数学语境中使用的术语——都将带来一代最低能量的弦振动模式。因而，有3个洞的卡拉比丘形可用来解释表12.1中基本粒子重复出现的三代结构。事实上，人们发现了很多的这种有3个洞的卡拉比丘形。在这些卡拉比丘形中，人们进一步挑出能给出正确的信使粒子数目、正确的电荷，以及正确的核力性质以匹配表12.1和表12.2中的粒子的卡拉比丘形。

这是一个鼓舞人心的结果，虽然还没有得到保证。在调和广义相对论以及量子力学的过程中，弦论可能已经达成了一个目标，只是我们发现几乎不可能达到另外一个同等重要的目标——解释已知物质和力的粒子的性质。那令人失望的可能性不会令研究人员退却。更进一步，计算出粒子的准确质量无疑是极具挑战性的。正如我们讨论过的，表12.1和表12.2中的粒子质量与弦最低能量的振动模式相差很大，足有千万亿倍。计算出这近乎于无限的差别超出了我们今日对弦论的理解。

事实上，我以及其他很多位弦论学家都在猜测，从弦论中得到表12.1和表12.2中的粒子的方式可能与从标准模型中得到这些粒子的方式非常相似。回想一下第9章，在标准模型中，整个空间中的希格斯场都取非零值，一个粒子所具有的质量的大小取决于当它试图从希格斯海中脱逃而出时感受到的拉力有多大。在弦论中很可能也有类似的机制。如果有数目巨大的弦在整个空间中按照相同的方式振动，那它们就会成为一种均匀的背景，而提供这种背景的意图与目的同希格斯海别无二致。最初没有质量的弦的振动模式，将通过在弦论版本的希格斯海中的运动和振动时感受到的拉力获得小的非零质量。

但需要注意的是，在标准模型中，给定粒子所感受到的拉力——也就是它所获得的质量——是由实验测量确定的并被作为输入参数放

到理论中。而在弦论中，这种拉力——也就是弦的振动模式的质量——可追溯到弦之间的相互作用（因为希格斯场也是弦的振动模式），因而是可计算的。弦论，至少在理论上，允许通过理论本身给出所有粒子的性质。

不过还没有人能实现这种构想，但必须强调的是，弦论仍在不断完善之中。一直以来，研究人员都希望完全搞清楚这一方法在统一方面的巨大潜力。这一动机如此强烈因为其潜在的回报实在太过丰厚。通过不断努力以及那么一点运气，弦论很可能会在某一天解释清楚基本粒子的性质，并且解释清楚宇宙为什么是这个样子。

弦论中的宇宙结构

尽管弦论的很多方面还不在我们的理解范围内，它已经为我们展现了很多奇妙的新景象。最令人吃惊的是，在填补广义相对论与量子力学间的鸿沟的过程中，弦论向我们揭示了基本层面的宇宙可能存在着远多于我们直观感受到的维度，而这些额外的维度很可能是探索宇宙最神秘之处的关键。而且，弦论告诉我们，熟悉的空间和时间概念不能被推广到亚普朗克尺度，这就意味着我们目前理解的空间和时间很可能只是某种我们现今尚未明晰的基本概念的近似。

在宇宙初创时期，时空结构的这些特点——今天的我们只能借助数学来探讨——可能是很明显的。早期，我们熟悉的3个维度也非常小时，我们现今在弦论中将其区别为大维度和蜷曲维度的这些空间维度，可能只有很小的区别，甚至根本没有区别。今天，这些维度在尺度上的不同，可能来自于宇宙演化。而宇宙演化，可能通过某种我们尚未理解的机制，挑出了3个特殊的空间维度，并将我们前面讨论过的长达140亿年的膨胀任务交给了这3个特殊的维度。沿着时间之箭进一步回退，整个可观测宇宙都缩小到亚普朗克范围，宇宙变成了图10.6中的模糊地带。而在这个模糊地带中，熟悉的空间和时间都产自于更加基本的实体，而搞清楚这些实体，不管它们到底是什么，正是

科学家们目前为止奋斗的研究工作。

要想进一步理解原初宇宙，乃至空间、时间的起源，以及时间之箭，我们必须使我们用来理解弦论——一个不久之前还看起来遥不可及的目标——的理论工具变得大大的锋利起来。现在我们已经看到，随着M理论的发展，进展已经超出了哪怕是最乐观的理论学家的最乐观估计。

## 第13章　膜上的宇宙

关于M理论中空间与时间的思索

在所有的科学发现中，弦论的发展轨迹最为崎岖。即使在其诞生30年后的今天，大部分的弦论学家仍然相信人们并没能对一个根本问题给予完备的回答，这个问题就是：什么是弦论？关于弦论，我们已经了解了很多。我们知道其根本特性，知道其主要的成就，知道它所给出的许诺，也知道它所面临的挑战；我们甚至还能使用弦论的方程来详细的计算出很多情形下弦的行为及其相互作用。但是大部分的弦论研究者仍然感觉到我们还缺乏那种人们在其他著名的科学成就中所拥有的核心原理。狭义相对论将光速视为常量；广义相对论拥有等价原理；量子力学拥有不确定原理。弦论学家们也在苦苦寻觅这样一种能抓住理论本质的原理来将弦论完善起来。

在很大程度上，这种缺陷的存在是因为弦论是零碎发展起来的，而不是基于某种深刻的洞察力。弦论的目标——将所有的力与所有的物质统一到量子力学的框架之下——是要尽可能的具有普适性，但是理论本身的演化却明显的散碎。弦论在30多年前偶然的诞生于人间之后，一些理论学家通过研究这些弦论的方程揭示了其关键性质，而另一些理论学家则通过研究那些关键的性质而发现了其暗含的深刻意义，弦论就是在这样的修修补补的过程中发展起来的。

弦论学家就像在其偶然绊倒的地方挖出了太空船的原始人一样。敲敲弹弹，原始人也可以慢慢的对太空船的操作方法有所感觉，逐渐发现所有的按钮和杆柄按一定的方式配合使用就可以操纵太空船。弦论学家们也有类似的感觉。多年的研究成果彼此吻合趋同。这样的事实给弦论学家带来了一种信心：弦论正在接近某一强大自洽的理论框架——虽然还不完善，但其最终必将以难以匹敌的清晰性和包容性揭示出大自然的内在规律。

近年来，被冠以第二次超弦革命的理论发展极好地体现了这一点。所谓的第二次超弦革命硕果累累，这次革命展示了盘绕于空间结构的隐藏维度，开启了新的实验检验弦论的可能性，提出我们的宇宙可能是众多可能性中的一种，发现在下一代高能加速器上可能创造出黑洞，还提出了新颖的宇宙学理论，在这一理论中，时间及其箭头，就像土星那优雅的光环一样，彼此缠绕。

第二次超弦革命

我在下面将会讲一个令人尴尬的弦论细节，读过我前一本书《宇宙的琴弦》的读者可能会记得，在过去的30年间，人们发展出的不是一个，而是5个截然不同的弦论版本。这些版本的名字无关紧要，我们将其分别称之为I型弦论，IIA型弦论，IIB型弦论，杂化O型弦论以及杂化E型弦论。所有的这些弦论都具有上一章介绍过的那些本质特征——其基本组成都是振动的弦——而且，正如20世纪70年代和80年代的有关计算所表明的那样，每一种弦论都需要6个额外的维度；但是，深入的分析表明，这5种弦论有着重大的区别。比如，I型弦论中的弦是上章讨论过的环，也就是所谓的闭弦，但是不同于其他弦论之处在于，这一理论中也有开弦——拥有自由的两端的弦。而且，计算告诉我们，不同版本的理论中，弦的振动模式以及每种模式彼此相互作用及影响的方式也各不相同。

弦论学家最想看到的当然是在未来的某一天，在与实验数据仔细

对比后，人们可以将这5个不同版本中的4个放到纸篓里。坦率地讲，弦论中存在着5个不同的版本毕竟不是一件舒服的事。统一之梦的关键在于科学家们会被带到一个关于宇宙的独一无二的理论面前。如果研究人员只需建立一个理论框架就可以将量子力学和广义相对论统一起来，那么物理学家们便到了真正的天堂。如果是这样，即使没有实验数据，人们仍然有理由相信所有的一切都是可靠的。毕竟，已经有大量实验证据支持量子力学和广义相对论。看起来很明显，统治宇宙的自然法则必须彼此兼容。如果存在一个独一无二的理论，可以在数学上自洽的将拥有坚实实验基础的20世纪两大物理学支柱统一起来，那么这个理论的合理性是有强有力——虽然是间接的——证据支持的。

可弦论却有5个表面类似但细节不同的版本，这使得弦论的独一性受到质疑。即使某些乐天派会辩称未来的某一天实验将从中挑出独一无二的理论，我们仍然会为存在额外的4个自洽的理论而感到恼火。难道另外的4个仅仅是数学上的巧合吗？它们对现实的物理世界是否有某种意义呢？它们的存在会不会只是冰山一角，未来的某一天会不会有聪明的理论学家发现实际上有5个额外的版本，或者6个、7个，甚至无限多个数学上自洽的不同弦论方案存在？

20世纪80年代末90年代初，很多物理学家热衷于探索这种或那种弦论，5个版本之谜并不是一个研究者日常每天关注的问题。相反，在人们的信念中，它只能算是在不久的将来随着对各种弦论的认识都达到一定的高度后将会自动得到解决的众多问题中的一个。

时间到了1995年春天，在几乎没有任何征兆的情况下，这些谨慎的愿望一下子被超越了。在数位弦论学家（包括克里斯·赫尔，保罗·唐森，艾索科·森，麦克尔·达弗，约翰·施瓦茨，以及其他一些人）工作的基础之上，爱德华·威滕——20年间一直是最有声望的弦论学家——发现了一种隐藏的一致性，从而将5种弦论一下子统一了起来。威滕证明，这5种弦论并不是彼此无关，而是数学上用来分析某种独

一理论的5种不同方式。正如将一本书翻译成5种不同的语言后，在不同语种的读者眼中，就是5种不同的书。5种弦论之所以不同，只是因为威滕还没能为这5种弦论编写好彼此之间翻译的字典。但是一旦编写完成，这样的字典就会明确告诉我们——正如人们可以从一个单独的母本得到5种不同的译本——5种弦论体系通过一个单独的母理论联系起来。这一统一的母理论可暂且称为M理论，M可以是很多意思——主体（Master）？宏伟（Majestic）？母（Mother）？魔幻（Magic）？神秘（Mystery）？矩阵（MatriX）？——只等世界范围内孜孜以求的研究者们最终完成威滕深刻的洞察力发现的这一理论，我们便可以给它一个真正的名字了。

这一革命般的发现真是令人欣慰的重大飞跃。威滕用这一领域中最为人称道的几篇论文（以及随后皮特·哈罗瓦的重要工作）证明，弦论是一个单独的理论。弦论学家们再也不用为挑选候选者而尴尬了，爱因斯坦所追寻的统一理论一度因为有5种不同的版本而失去统一性，现在这种情况一去不返了。而且，一个统一理论能达到的最高程度上的统一就是统一于其自身。通过威滕的工作，每种独立的弦论所实现的统一性被扩充到整个弦论框架。

图13.1刻画的是威滕提出其发现前后5种弦论的状况，这一总结值得我们记住。正如图片所示的那样，本质上，M理论并不是一种新的方法；云消雾散之后，人们发现M理论承诺的是比任何一种单独的弦论所能提供的更加精确更加完备的物理定律体系。M理论将5种弦论联系起来了，5种弦论的每一个只不过是更加伟大的理论体系的一个部分。

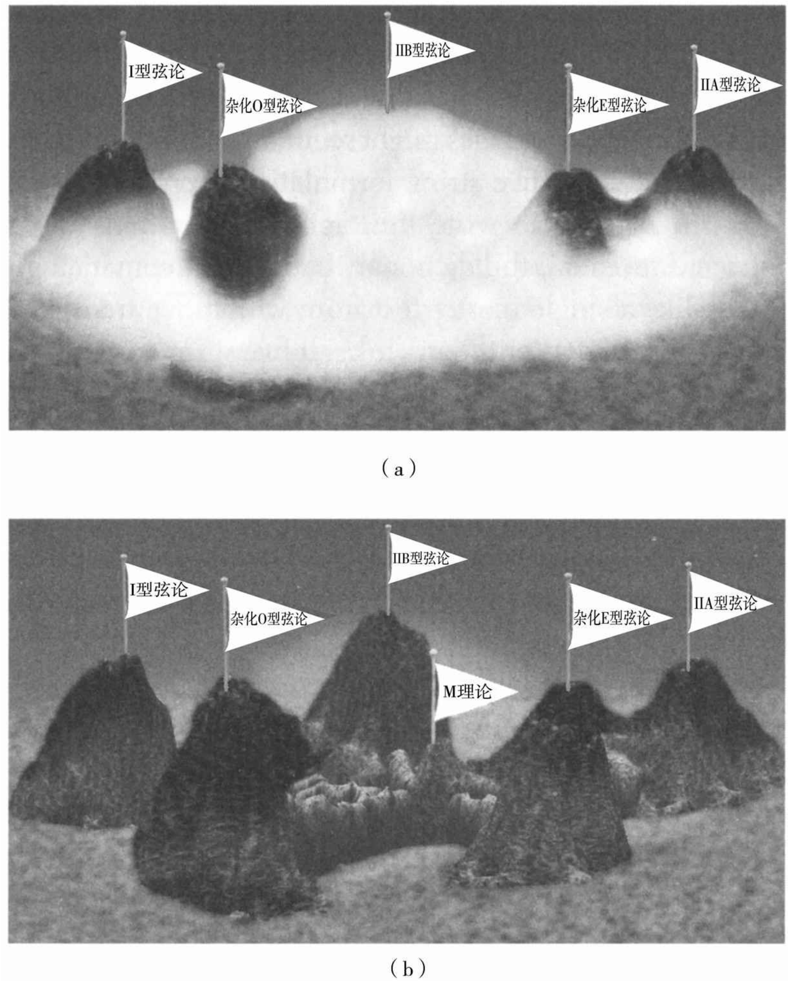



图13.1　（a）：1995年之前弦论研究的示意图。（b）：M理论所揭示的统一示意图。

## 翻译的力量

图13.1示意性的说明了威滕发现的核心内容，但是这种表述方式可能使你感觉非常的具有技巧性。在威滕取得突破之前，研究者们认为有5种不同版本的弦论；在那之后，情况发生了变化。但是，要是人们从来都不知道有5种不同的弦论的话，那么最聪明的弦论学家证

明了这5种弦论并不是全然无关的又有什么意义呢？换句话说，我们为什么将威滕的发现视作一场革命，而不仅仅是纠正了之前错误概念的一个普通观点呢？

现在我们就来回答这个问题。过去的几十年，弦论学家们一直被一个数学问题困扰。描述5种弦论中的任何一种的精确方程都极其难于处理分析，因此理论学家们常常被迫求助容易解决得多的近似方程。人们有理由相信，近似方程在很多情况下都会给出接近于真实方程结果的答案，但是近似——就像翻译一样——总会丢点什么。正因为如此，一些近似方程力所不及的关键问题显而易见地影响了理论的进展。

对于文字翻译中的不当之处，读者会有所知觉并加以修缮。有些语言功底很好的读者甚至可以直接反推原文。但是，弦论学家们却没有这样的机会。纯粹依靠威滕等人编写出来的字典的自洽性，我们就有很充分的理由相信5种弦论只不过是一种母理论——M理论——的5种不同描述，但是，弦论的研究者们还不能完善的理解这种理论上的联系。过去的几年间，我们在M理论的研究方面已经取得了一些进展，但是离最终准确完整的理解M理论仍有很长的路要走。在弦论的研究方面，我们好像得到了一个看似即将发现的母本的5种译稿。

对于很多读者来说，他们可能既没有原文（就像弦论），也不是很了解原文所用的语言，他们希望的是参考几种翻译成他们所熟知的语言的译文，以便更准确地了解原文。要是某些段落在不同语言的译本上彼此一致，读者们就会对这些译文有信心；要是彼此不一致，那就意味着翻译上的不准确或者不同的诠释。就是这样，威滕发现了5种不同的弦论只不过是深层理论的不同译本。事实上，威滕的发现开创了一条极其强大的战线，通过翻译的类比我们可以很好地理解这一点。

我们完全可以想象有这样一种母本，其中有数不清的双关语，押

韵或不押韵的诗句，特定文化中的笑话，这样的文字很难完整优美的从5种不同的译本中的单独一种翻译回原文。某些段落可能很容易翻译成斯瓦希里语，但其他段落则可能完全不适合这种语言；后面的这些段落可能非常适合翻译成因纽特语，但是其他的部分就可能翻译得含混不清。梵语或许能捕获某些深奥章节的妙意，但是另外一些章节，可能所有的5种语言全部束手无策，只有写成原文的语言才最适合表达那个意思。5种弦论的研究就处于这样一种状况。理论学家们发现，对于特定的问题，5种弦论中的某一种可能会给出明晰的物理描述，但是另外几种就会在数学上太过复杂以至于难以应用。威滕的发现其威力正在于此。在他取得重大突破之前，弦论的研究者们要是遇到了非常难于处理的方程可能就会卡在那里。而威滕的发现告诉我们每一个这样的问题都有另外4个译本——4种数学形式——很多时候其中的某一个并不是那么难于应付。因此，这本用于不同理论之间互译的字典常常可以用来将困难的问题翻译成相对简单的问题。

不过这并不总是非常简单。就像对于某些段落的翻译，所有的译本给出的翻译可能都不能令人满意，同样的，弦论中的某些问题在所有的5个理论中也可能都非常的难于理解。这个时候，我们就只好参考原文了；也就是说，只有完全理解了难懂的M理论我们才能取得进展。不过，这并不妨碍威滕字典的巨大威力，在很多情况下，威滕的字典都可以作为我们分析弦论的新的强有力工具。

正如一段复杂文字的每一种翻译都为一个重要的目标而存在，每一种弦论也是如此。将我们从每一种单独的弦论中所学到的知识整合在一起，我们就可以回答那些单独的一个弦论无法回答的问题，发现那些超越了单独的每个弦论的特征。威滕的发现为理论学家们带来了5倍的强大火力来推进弦论研究的战线。正因为如此，我们才说威滕的发现引发了一场革命。

**11个维度**

新发现的强有力的工具为我们的弦论研究带来了什么样的成果呢？成果是大量的。这里，我会将注意力集中到对空间和时间的故事具有最重大影响的几个成果上。

首先，威滕工作最重要的发现是人们在20世纪70～80年代使用近似方程所得到的结论——宇宙必须有9个空间维度——实际上差了一个数。威滕的分析表明，根据M理论，宇宙应该有10个空间维度；也就是说，时空是十一维的。就像卡鲁扎发现一个五维的时空可以为电磁力和引力的统一提供理论框架，以及弦论学家们发现具有十维时空的宇宙可以为量子力学与广义相对论的统一提供理论框架；威滕发现具有十一维时空的宇宙可以用来统一所有的弦论。这就像在平地上看远处的村庄，5个不同的村子看起来彼此完全分开，但要是站在高山上的话——利用这个额外的垂直维度——人们就会发现这5个村子实际上是通过小路和大道联系起来的。在威滕的分析中引申出来的额外空间维度是在5种弦论之间找到联系的关键。

尽管威滕通过额外维度获得重大发现是符合达到统一理论的历史轨迹的，但是当他在1995年的年度国际弦论会议上抛出他的结果时，整个研究领域内还是能感到强烈的震动。包括我在内的大部分研究人员长久以来一直研究近似方程，大家都认为所做过的分析已经最终回答了维度的数目这一问题。但是威滕，还是带来了令我们惊奇不已的东西。

威滕证明，之前所有的分析都采用了一种数学上的简化，这一简化等价于假设有一个迄今尚未认清的维度极其的小，比所有其他的维度都要小很多。事实上，这一维度实在是太小了，小到所有研究者使用的近似方程都没有能力发现这样一种极小维度存在的数学线索。因而所有的人都认为弦论只有9个空间维度。但是，依靠统一的M理论的新视角，威滕可以超越近似方程，更加细致地探索问题，从而发现有一个空间维度长久以来一直为人们所忽略。最终，威滕证明10多年来理论学家们研究而得的5个十维弦论体系实际上是一个单独的深层

十一维理论的5种不同的近似描述。

你或许想问这个未料到的发现是否会使以前人们在弦论方面的工作变得一无是处。大体上讲，并不是这样。虽然新发现的第10个空间维度为理论带来了一个未曾预期的特性，但如果M理论真的是正确的话，那么第10个空间维度就应该比其他的空间维度小很多——就像很长一段时间以来人们在毫不知情的情况下假定的那样——人们之前所做的工作仍有其用武之地。不过，考虑到已知的方程仍然不能明确额外维度的大小和形状，弦论学家们在过去的几年花了很多的汗水来探索没那么小的第10个空间维度的可能性。其他的姑且不提，理论学家们努力研究所得到的大量结果将图13.1中关于M理论统一威力的示意性说明放在了坚实的数学基础之上。

我猜从十维升级到十一维——姑且忽略其对弦论或M理论数学结构的重要性——并不会明显的改变你对这个理论的想象。对于外行人来说，想象7个蜷曲起来的维度实在不比想象蜷曲起来的6个维度难多少。

但是与第二次超弦革命紧密联系的另一件事的确改变了弦论的直观图象。几位科学家——威滕、达弗、赫尔、唐森，以及其他几位——的集体智慧使人们相信，弦论并不仅仅是弦的理论。

膜

你在上一章一定遇到一个很自然的问题——为什么是弦？为什么一维的东西这么特别？在调和量子力学与广义相对论的过程中，我们发现关键之处在于基本组分要有大小，要是弦而不是点。两维的东西也可以有大小呀，比如微型的碟片或飞盘；又或者是三维的东西，比如棒球或土块之类的。甚至，既然理论可以有非常多的维度，那么基本组分完全可以是更高维的某种东西。那么，这些东西为什么不能在基本理论中扮演任何角色呢？

20世纪80年代与90年代早期，很多弦论学家似乎找到有一定说服力的回答。他们认为人们已经多次尝试利用高维的块状物作为基本组分来构建基本理论，这其中包括20世纪物理学的偶像人物，比如沃纳·海森堡和保罗·狄拉克。他们以及其后的很多研究工作表明，利用很小的块作为基本组分建构理论极其困难，这样的理论常常很难满足基本的物理学要求——比如，保证量子力学概率在0和1之间（负的和大于1的概率毫无意义），以及不能超光速通信。至于点粒子，始于20世纪20年代的半个世纪的研究表明，所有的这些要求都可以满足（当然，引力要暂且忽略）。到了20世纪80年代，施瓦茨、谢尔克、格林以及其他人10多年的探索使人们惊奇的发现，一维的弦也可以满足所有的那些要求（且必然包括引力）。但是，进一步将基本组分视为两维或更高空间维度的客体则几乎不可能。究其原因，则是方程所要考虑到的对称性在基本组分为一维客体（弦）的时候达到最大，之后迅速衰减。这里方程中的对称性比第8章中讲到的还要抽象（研究弦或更高维的东西的运动时，我们要一会儿靠近点，一会儿离远点，因而我们考虑的对称性就与突然改变我们的观测分辨率时方程如何变化有关）。这些转变对构建物理上合理的方程组非常关键，对于比弦更高维度的研究对象，这些所需要的对称性消失了。[1]

因此，当威滕的论文以及其后的大量研究工作[2]表明弦论及其所属的M理论中可以包含弦之外的基本组分的时候，弦论学家们再次被震惊了。研究显示理论中可以有两位的客体，很自然的，我们将其称之为膜（membrane，M理论中M的另一种可能解释），或者——为了系统的命名其高维兄弟——我们将其称之为2-膜。当然还有具有3个空间维度的客体，我们将其称为3-膜。虽然很难形象化的想象出来，但是研究表明还有具有P个空间维度的客体，其中P可以为小于10的整数，它被称为——毫无悬念——P-膜。因此，弦只是弦论中的一种组分，而不是唯一的组分。

这些组分之所以能够逃脱之前的理论研究，其原因与第十维度能

够逃脱一样：近似方程太过粗糙以至于没能发现这些家伙。在弦论学家们用数学研究的理论框架内，所有的P-膜都要比弦重很多。一个东西的质量越大，产生它所需要的能量就越多。近似方程的局限之处在于——方程本身固有的局限之处，所有的理论学家们都很清楚——当所描述的实体和过程与越来越多的能量有关的时候，方程本身的精确性就越来越差。当达到与P-膜有关的极端能量时，近似方程就会失准，从而使膜潜伏于黑暗之中，这就是人们几十年都未能在数学上发现膜的原因。但是利用M理论提供的各种新方法，研究者们可以绕过之前的一些技术壁垒，从而依靠完整的数学视角，他们发现了盛装列队的高维组分。[3]

弦论中除了弦还可以有其他维度的组分存在，这一点并不比第十维的发现更易使早前的工作无效或过时。研究表明，要是高维的膜比弦重很多的话——就像在早前的工作中无意识的假定的那样——它们对大部分的理论计算几乎没什么影响。不过，就像第十维度并不是非得比其他9个维度小很多一样，高维的膜也不是必须非常的重。在很多理论假定的情况下，高维膜可以同最低质量的弦的振动模式有相同的质量，这时，膜就会对物理世界有重要的影响。比如，在我与安德鲁·斯特劳明格和戴维·莫里森合作的工作中，我们证明了膜可以将自己绕在卡拉比丘流形的球面上，就像绕在柚子外面的用于真空封闭的塑料薄膜一样；要是那个流形收缩，绕在上面的膜也会跟着收缩，导致其质量减小。我们证明，这一质量减小会使空间完全坍塌撕裂——空间把自己撕破了——而紧绕的膜会保证不发生灾难性的物理后果。我在《宇宙的琴弦》一书中详细讨论了这一问题，在第15章讨论时间旅行的时候我们还会再简要的讨论一下这个问题，这里就不作详细阐述了。不过，从这个小插曲上我们可以清楚的看到高维的膜是如何对弦论的物理世界产生影响的。

回到我们当前的焦点所在，根据弦论或M理论，膜还可以以另外一种深奥的方式影响我们对于宇宙的认识。这宏伟辽阔的宇宙——我

们所熟知的全部时空——可能本身就是一张巨大的膜。我们的宇宙可能是一个膜世界。

膜世界

检验弦论的正确性是一个巨大的挑战，因为弦实在太小了。但请别忘了决定弦的尺寸的物理。引力的信使粒子——引力子——是弦最低能量的振动模式，引力子交换的引力强度正比于弦的尺寸。引力如此之弱，因而弦的长度必定极小；计算表明，要使弦的引力振动模式交换的引力达到所观测到的大小，弦的长度就只能在普朗克长度的100倍左右。

从这个解释中，我们可以看出很高能量的弦并不是非得很小，只要它不和引力子（引力子是低能零质量的振动模式）直接联系就行。事实上，随着弦获得越来越多的能量，其振动也会变得越来越猛烈。但是，一旦突破某个特别的点，弦额外获得的能量就会有其他的作用：这些能量会使弦变长；而且，对于最终能有多长，并没有任何限制。因而，只要你不停地将能量注入弦中，它甚至可以长到宏观量级。以今天的技术水平，我们不可能做到这一点。但是，在大爆炸之后的极热、超高能的情况下，很长的弦可以产生出来。如果其中的某些长弦直到今天还存在的话，它们可能横跨天际。虽然可能性很小，而且更可能的情况是这样的长弦可能也非常的小，但是我们的确有可能在获自太空的数据中发现长弦存在的蛛丝马迹。或许某一天我们可以从天文学的观测中检验弦论。

高维度的P-膜也不是非得很小，考虑到P-膜比弦的维度还要多，定性上讲，存在着很多新的可能性。当我们试图描画一根长——或许是无限长——弦的时候，我们就会想象出一根存在于我们生活于其中的三维空间的很长的一维客体。就像一根长到我们视野的边际的电线一样。类似的，如果我们要描画一张很大的——或许是无限大——2-膜，我们就会想象出一张存在于普通的三维空间中的二维表面。我想

不出什么好的现实类比。或许我们可以想象一家拥有超大屏幕的汽车影院，那超大的屏幕非常的薄，又宽又高，挡住我们能看到的所有地方，这样的屏幕或许会帮我们想象一下2-膜。但要是我们试图理解一下3-膜的话，我们就会发现自己遇到了新问题。3-膜有3个空间维度，因而，如果它非常大的话——或许可以是无限大——就会填满整个的三维空间。1-膜和2-膜，就好比是电线和屏幕，它们存在于我们的三维空间之内，而3-膜占据的是我们整个的三维空间。

于是就有了一种非常吸引人的可能性。有没有可能我们此刻是生活在3-膜上呢？就像白雪公主，她的世界存在于二维的屏幕——一张2-膜——而二维的屏幕又存在于三维的宇宙（影院的三维空间）。我们所知晓的一切有没有可能是存在于一张更高维度的屏幕——一张3-膜——而这张更高维度的屏幕又是存在于弦论或M理论的高维的宇宙呢？牛顿、莱布尼茨、马赫、爱因斯坦这些人称之为三维空间的东西有没有可能是弦论或M理论中某种特别的实体呢？又或者用相对论的语言说，闵科夫斯基和爱因斯坦提出的四维时空可不可能是一张3-膜随时间演化时留下的轨迹呢？简而言之，我们知晓的宇宙可不可能是一张膜呢？[4]

我们生活在一张膜上这一可能性——所谓的膜世界方案——是弦论或M理论的最新演绎。我们将会看到，它使我们可以以一种新颖的方式思考弦论或M理论，这些想法枝繁叶茂。这里物理上的关键之处在于，膜很像是宇宙的威扣；[68]在某种特别的方式下，它们可以非常的黏。我们马上来讨论一下。

### 黏黏的膜与振动的弦

引入“M理论”这一术语的动机之一就是我们已经认识到“弦论”这一术语只能展现出理论的某一方面。早在精炼的分析发现了高维膜的十几年前，理论研究就已经展示了弦的存在，所以“弦论”这一术语有其历史因素。不过，即使M理论允许各种各样维度的客体彼此平等，

弦也要在我们的公式体系中扮演关键角色。原因之一很容易说清。正如20世纪70年代的研究者们不自觉地加以运用的那样，当所有的P-膜远重于弦的时候，它们都可以被忽略。但是，弦之所以非常特别还有更加具有普遍意义的原因。

1995年，威滕宣布他的突破之后不久，加利福尼亚大学圣芭芭拉分校的乔·波金斯基想到了几年前他与罗伯特·利和戴锦（音译）合作的论文，在那篇论文中，波金斯基发现了弦论有趣但相当晦涩的一个性质。波金斯基的动机和推导相当的技术化，我们不打算讨论其细节，只关心其结论。波金斯基发现，在某些情况下，开弦——还记得吗？那些两端松散的弦——的端点并不能完全自由的移动。就像用绳串起来的珠子，虽然可以自由移动，但是必须沿着绳；又像小时候玩的弹球游戏一样，球虽然可以自由移动，但是必须按照弹球桌表面的形状移动；开弦的端点也是这样，在空间的特点、形状或轮廓的限制下可以自由移动。在考虑到弦本身还可以自由振动后，波金斯基及其合作者证明弦的端点在某些区域会变得“很黏”或者说被“套牢”。

在某些情况下，这样的区域可以是一维的，此时弦的端点就像是用绳子串起来的珠子，而弦本身就像是那根绳子。又或者在另外的情况下，这样的区域可以是二维的，此时弦的端点就像是被一根绳子连起来的两颗弹球，它们只能在弹球桌上滚动。还有其他的情况，这样的区域可以是三维、四维，或者小于10的任意空间维数。波金斯基以及皮特·哈罗瓦和麦克尔·格林证明的这些结果，解决了在比较开弦和闭弦时长久存在的一个问题。不过他们的工作在最初的几年并没有引起人们足够的重视。[5]直到1995年10月，在威滕新发现的启发之下，波金斯基重新审视了早前的这些工作之后，情况才发生了改变。

在波金斯基的论文中，有一个问题并没有得到完全的解决，或许你在读上一段时曾意识到这个问题：如果开弦的端点在空间的某些区域可以被粘住，那么粘住它们的又是什么？绳子与弹球桌都可以脱离被限制于其中的珠子和弹球而独立存在。那么开弦的端点被限于其中

的空间区域呢？这些空间区域中是否充满了弦论中的某些独立且基本的组分？是否正是这些家伙无情地抓住了开弦的端点？1995年之前，弦论还只是弦的理论的时候，人们没法找到能做这件事的候选者。但是在威滕的重大发现以及随之而来的大量成果涌现出来之后，波金斯基想到了问题的答案：如果开弦的端点被限制在空间中的某些P-维区域上运动，那么这样的区域必须被P-膜占据。69波金斯基的计算表明新发现的P-膜正有那种将开弦的端点牢牢抓住的能力，从而使开弦的端点只能在充满P-膜的区域内运动。

为了更好地理解这些内容，我们来看一下图13.2。在图（a）中，我们可以看到一对2-膜，很多开弦在其上来来回回振动不息，这些开弦的端点都被限制在各自运动的膜上。高维膜的情况非常类似，只是难以画出。开弦的端点可以在P-膜上自由的运动，但是不能离开膜。当它们想离开膜的时候，就会发现膜是黏到难以想象的东西。开弦的两个端点也可以分别附着于两个不同的P-膜，这两个P-膜的维数可以相同[图13.2（b）]，也可以不同[图13.2（c）]。

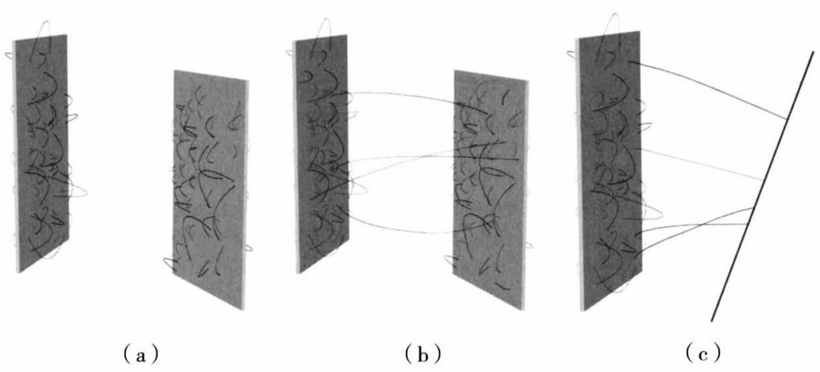



图13.2　（a）：端点附着于二维膜（或者说2-膜）的开弦。（b）：两个端点分别附着于两个2-膜的开弦。（C）：两个端点分别附着于一个2-膜和一个1-膜的开弦。

紧跟着威滕对各种弦论之间联系的发现，波金斯基发表了他的第二次弦论革命宣言。尽管20世纪理论物理学界中一些最聪明的头脑努

力奋斗但却没能建立一个包含大于点（零维）和弦（一维）的维度基本组分的理论体系，但是威滕和波金斯基的工作，再配以很多当代顶级理论学家的深刻洞察力，物理学家找到了前进之路。这些物理学家并不仅仅建立了包含高维组分的弦论或M理论，波金斯基的洞察力还为理论上精细分析其中的物理性质提供了方法（当然前提是这些东西要真的存在）。波金斯基主张，在很大程度上，膜的性质是由端点附着在膜上的振动着的开弦的性质决定的。这有点类似于我们通过抚摸地毯上的绒毛来了解地毯的质地——那些绒毛的另一端就固定在地毯上——我们也可以通过研究一端附着在膜上的弦的性质来认识膜。

这一结果非常重要。它证明了这几十年来为了研究一维对象——弦——而制造的精妙数学工具也可以用于高维对象的研究，这个高维研究对象就是P-膜。而妙不可言的是，波金斯基发现对高维对象的研究可以在很大程度上约化为对理论上更加熟悉的弦的研究，尽管弦也还是理论假设。在这种意义下，弦的特别之处就显现出来了。如果你懂得弦的性质，那么你就在一条通往理解P-膜的漫长道路上了。

有了这些领悟，让我们再回到膜世界方案——我们全都生活在3-膜上的这种可能性。

若我们的宇宙就是一张膜

如果我们生活在一张3-膜上——如果我们的四维时空只是一段3-膜在时间的长河中掠过的历史——那么我们就可以用全新眼光审视时空究竟是不是某种实物这一古老的问题了。我们所熟知的四维时空应是来自于弦论或M理论中的一种实体——3-膜，而不是某种模糊的抽象概念。按这种理念，我们的四维时空的实在性就会与电子或夸克的实在性一样（当然了，虽然你知道我们所直接感受到的时空舞台具有明显的实在性，可你还是会追问弦与膜存在于其中的更大时空——弦论或M理论的十一维时空——本身是否仍是实体）。但如果我们所知晓的宇宙真的是一张3-膜，那我们岂不很轻易就会知道有某种东西

——3-膜——弥漫在我们周围？

不过嘛，我们已经学过的现代物理告诉我们，虽然在我们的身边的确弥漫着很多东西——希格斯海，满是暗能量的空间，无数的量子场涨落——但仅凭人们的直觉，所有的这些我们都没法感受到。因此，知道了弦论或M理论在“真”空的不可见事物单中又加了一项，实在没什么值得惊讶的。但让我们先别自以为是。上面提到的每一种可能，其对物理的影响，以及我们应如何证明其存在，我们都非常清楚。事实上，上面提到的三者之二——暗能量与量子涨落——我们已经看到有足够多的证据证明其存在了；至于希格斯场，当前或未来的对撞机实验将给予答案。那么，我们是否生活在3-膜之中这个问题是不是也与之类似呢？如果膜世界方案正确的话，我们为什么没有看见3-膜呢？我们究竟该怎样才能证明其存在呢？

这一问题的答案将清楚地告诉我们，膜世界方案中弦论或M理论的物理推论与早前的“免膜”方案（有时候人们也会将之亲切地称为无膜方案）到底有多么大的差异。让我们来看一个重要的例子，光的运动——光子的运动。弦论中的光子，如你所知，是弦的一种特殊振动模式。但是数学分析表明，在膜世界方案中，只有开弦的振动才能产生光子，而闭弦则不能，这就是膜世界方案与先前的理论的一个重大区别。虽然开弦的端点只能在3-膜上移动，但在3-膜上的运动却是完全自由的。这意味着光子（开弦的振动产生的光子）在我们的3-膜中的运动既不会受到任何限制也不会遇到任何障碍。而这会使膜看起来完全透明——完全不可见——于是我们根本看不出自己实际上浸泡其中。

而同样重要的是，因为开弦的端点不能离开膜，所以它们不能进入额外维度。正如电缆中的金属线决定着外面的绝缘皮的形状，弹球机中的球会被限制在一定的路线内，我们那黏黏的3-膜也会对其中的光子有所限制——光子只能在我们的三维空间中运动。因为光子为电磁场的信使粒子，因而对光子的限制也就意味着电磁力——光——只

能被囿于我们的三维世界，如图13.3所示（我们画出的只是二维示意图）。

认清这一点将为我们带来非常重要的结果。在前面，我们要求弦论或M理论中的额外维度要很紧的蜷曲起来。很明显，之所以有这种要求是因为我们看不到额外维度，因而我们必须想办法把它们藏起来。而把它们藏起来的办法中有一种就是令它们小于我们以及我们用来探测的器材。但现在让我们在膜世界方案中重新审视一遍这个问题。我们究竟怎样探测事物呢？这个嘛，当我们使用肉眼的时候，我们用的实际是电磁力；当我们使用电子显微镜之类的强大器材时，我们用的实际也是电磁力；当我们用原子对撞机时，我们用来探测超微观尺度的力中的一种还是电磁力。但如果电磁力被限制在我们的3-膜——三维空间中，那么，不论额外维度的尺寸有多大，仅有电磁力将无法探测到其存在。光子不能逃出我们的维度，进入额外维度中，再返回我们的眼睛或设备中以便我们能够探测到额外维度的存在，甚至在额外维度与我们熟悉的维度一样大的情况下也不行。

所以，如果我们真的生活在3-膜中，我们就有了另一种感受不到额外维度的解释了。额外维度并不是非得特别的小，它们也可以非常大。我们之所以看不到额外维，在于我们看额外维的方式。我们是用电磁力来看额外维度，而电磁力只能存在于我们的三维世界，无法进入额外维度中，就像睡莲叶子上的蚂蚁根本无从知晓叶子下面的深水，我们也可能漂浮在巨大而广阔的高维空间中，如图13.3（b）所示，但是电子力的特点——永远都无法逃离我们的三维世界——使我们没办法发现这一点。

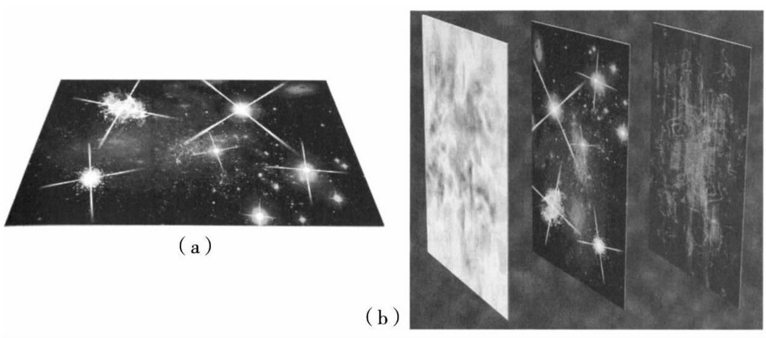


图13.3 （a）：在膜世界方案中，光子为端点固定在膜上的开弦，所以光不能离开膜本身。（b）：我们的膜世界可能漂浮在还有额外维度的更高维空间中，而那些额外维度是我们看不到的，因为光不能离开我们的膜。附近可能也漂浮着其他的膜。

好吧，你可能会说，电磁力只是大自然中的4种力中的一种，那其他3种呢？我们可以用它们来探测额外维度并发现其存在吗？对于强核力以及弱核力，答案还是：不行。在膜世界方案下，计算告诉我们这些力的信使粒子——胶子、W粒子与Z粒子——也是开弦的振动模式。所以它们像光子那样被囿于3-膜之中，有强核力与弱核力参与的过程同样无法与额外维度联系起来。对于物质粒子来说同样的结论依然成立。电子、夸克以及其他所有种类的粒子都是端点在3-膜上的开弦的振动模式。因而，在膜世界方案中，你、我以及我们能够看到的一切都被永远的拘禁在我们的3-膜中。把时间维度也算上的话，世间万物都困在我们的四维时空片中，其实应该说，几乎世间万物，因为对于引力来说，情况就有所不同。对膜世界方案做一番数学分析后我们会发现，引力子来自于闭弦的振动模式，这一点同我们之前讨论过的无膜方案中的情形是一样的。闭弦——没有端点的弦——并没被限制在膜上。闭弦自由自在，既可以在膜上运动，也可以离膜而去。所以，如果我们生活在膜上，我们并没有完全的隔绝于额外维度。通过引力，我们既可以影响额外维度也可以被额外维度影响。在这样的方案中，引力是我们能与三维之外的额外维度取得联系的唯一办法。

在我们通过引力与额外维度取得联系之前，我们难免会想知道，

额外维度究竟会有多大呢？这个问题非常有趣也非常尖锐，我们就来一起看看。

## 引力与大额外维

退回到1687年，牛顿提出了他的普适引力定律，在这个定律中，牛顿实际上对空间维数做了很强的限定。牛顿并没有仅仅用嘴说两个物体之间的引力会随着物体间距离的变大而变弱，他提出了一个公式，即平方反比率，这个公式准确地描述了两个物体间的距离发生变化时物体间引力的变化规律。根据牛顿公式，如果你将两个物体间的距离翻番，那两者之间的引力就会变为原来的四（22）分之一；如果你将两者间距离变为3倍，那引力就会变为原来的九（32）分之一；如果你将两者间距离变为4倍，那引力就会变为原来的十六（42）分之一；更为一般性的说，引力会按距离平方衰减。在过去几个世纪的大量实践中，牛顿这一公式始终有效。

但问题在于力为什么随物体间距离的平方变化呢？力为什么不是随物体间距离的3次方（这样的话你将距离翻番，引力就会变为原先的1/8）或4次方（你将距离翻番，引力就会变为原先的1/16）变化呢？或者更加简单些，两个物体之间的引力为什么不是就反比于（这样的话，你将距离翻番，引力就会变为之前的1/2）距离呢？这些问题的答案与空间维数直接相关。

要想看出这一点，可以想想两个物体吸收和发射的引力子数目是怎样依赖于间距的，也可以想想两个物体所体验到的时空曲率是怎样随着两者间距离的增加而减小的。不过我们可以用一个虽然过时但相对简单些的办法来思考这个问题，这会使我们快速而且直观的得到这一问题的正确答案。我们先画一张图[图13.4（a）]，就像图3.1示意的是条形磁铁所产生的磁场那样，我们这里的图用来示意一个有质量的物体——比方说太阳——产生的引力场。磁场线从北极出发，绕着条形磁铁收于南极；引力场线则不同，它是呈放射状向所有的方向发

射出去，直至无穷远。在给定位置处另一个物体——比方说绕太阳运动的卫星——所感受到的引力强度——正比于该位置处的场线密度。穿过卫星的场线数目越多，如图13.4（b）所示，则卫星感受到的引力就越大。

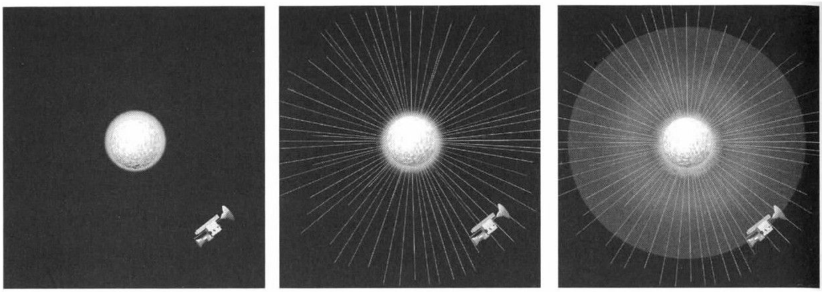



图13.4　图（a）中，太阳作用在物体（比如说卫星）上的引力，反比于其间距离的平方，因为太阳引力场线如图（b）所示的那样均匀延展，因而，到太阳距离为d处的场线密度反比于以d为半径的球面——如图（C）所示——的面积，而这个面积，根据几何知识，应正比于$d^2$。

现在我们可以解释一下牛顿的平方反比率的起源了。让我们一起想象一个以太阳为中心并将卫星包括在内的球，如图13.4（c）所示，其表面积——同任意一个三维空间中的球体的表面积一样——正比于其半径的平方，在这里就是太阳到卫星距离的平方。这就意味着穿越球面的场线的密度——总的场线数除以球的表面积——将随着太阳到卫星之间的距离的平方的增加而减小。如果你把太阳—卫星间的距离翻一番，同样数目的场线将平均分布于4倍大的表面积上，因而在这个距离上的引力就会变为之前的1/4。因而，牛顿的平方反比率引力公式实际上是三维空间中的球体的几何性质的反应。

与之相比，如果宇宙只有两个空间维度或者只有一个空间维度的话，牛顿引力公式该是什么样子呢？这个嘛，图13.5（a）给出的就是太阳与其环绕卫星的二维版本。因为圆周的长度正比于其半径的大小（而不是正比于其半径的平方），如果你将太阳与卫星之间的距离

翻一番，场线的密度就会减小为原来的二（而不是四）分之一，因而太阳的引力就变小为二（而不是四）分之一。如果宇宙只有两个空间维度，那么引力就会反比于距离，而并非是反比于距离的平方。

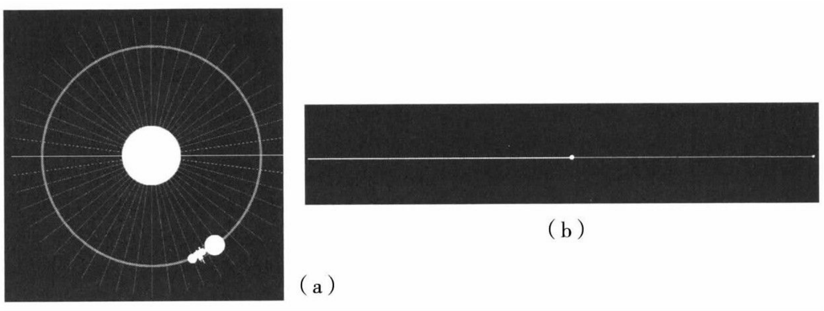



图13.5 （a）：在一个只有两个维度的宇宙中，引力反比于距离，因为引力线按环形均匀的延展，而环形的周长正比于其半径。（b）：在一个只有一个维度的宇宙中，引力线没有地方延展，所以，不论间隔多远，引力都是常数。

如果宇宙只有一个空间维度，如图13.5（b）所示，那么引力定律就会变得更加简单。引力场线根本就没有地方弥散，所以引力根本不会随着距离变大而变小。如果你将太阳和卫星之间的距离翻番（假设在这样的宇宙中存在着这样的卫星），你会发现穿过卫星的引力场线还是那么多，两者之间的引力一点变化都不会有。

虽然不能画出来，但是我们得知道，图13.4和13.5所示意的道理可以直接推广到四维、五维、六维乃至其他任意维度空间的宇宙。空间的维数越多，引力线延展开来的余地就越大。而引力线展开得越厉害，引力随距离变大而衰减的速度就越快。如果空间维数是4，牛顿定律就应该是立方反比率（距离翻番，引力变为之前的1/8）；如果空间维数是5，牛顿定律就应该是4次方反比率（距离翻番，引力变为之前的1/16）；如果空间维数是6，牛顿定律就应该是5次方反比率（距离翻番，引力变为之前的1/32）；对更高维的宇宙不过如是种种。

你可能会认为牛顿的平方反比率在解释大量的数据——从行星的运动到彗星的轨迹——上的成功证明我们生活在一个正好有3个空间维度的宇宙中。但是这个结论未免下得有些匆忙。我们知道，在天文学尺度上平方反比率的确被符合得很好；[6]而我们也知道，在地球尺度上平方反比率被符合得也很好，而且与我们看到的3个空间维度这一事实相一致。但是我们是否知道，平方反比率在更小的尺度上是否有效？引力的平方反比率是否被检验到多微观的尺度上了呢？实际上，实验学家对引力的平方反比率的验证只到毫米量级；如果两个物体的间隔有几个毫米大，那么实验数据告诉我们两者间的引力符合平方反比率的规律这一预言。但是再往下，继续探测更小尺度上的平方反比率就似乎是对实验技术的一大挑战了（量子效应以及引力非常微弱的特点使得实验非常复杂）。这个问题非常重要，因为对平方反比率的偏离将是存在额外维度的有力证据。

为了看清这一点，让我们用一个易于画出并分析的低维度玩具模型来说明问题。想象我们生活在一个只有一维空间的宇宙中——在这里，人们只能看到一个空间维度，而且几个世纪以来的实验证明物体间的引力绝不会因为物体间间距的变化而改变。在这个宇宙中，实验学家这些年来也忙着验证毫米尺度上的引力定律，没有得到过小于毫米量级的实验数据。进一步想象一下，这个宇宙，除了少数理论物理学家外无人知晓，实际上还有一个蜷曲起来的空间维度，如此则这个宇宙的形状就像菲利普·佩蒂特[70]的钢丝的表面，如图12.5所示。这会对未来更加精确的引力实验有什么影响呢？我们可以通过图13.6来推测答案。当两个很小的物体彼此靠近时——近到两者之间的距离比蜷曲维度的周长还要小时——空间的二维性就会立即显现出来，因为在这个尺度上引力场线本来就可以延展开来[如图13.6（a）所示]。因而，引力就不再是与距离无关了，而是反比于两个物体之间的距离。

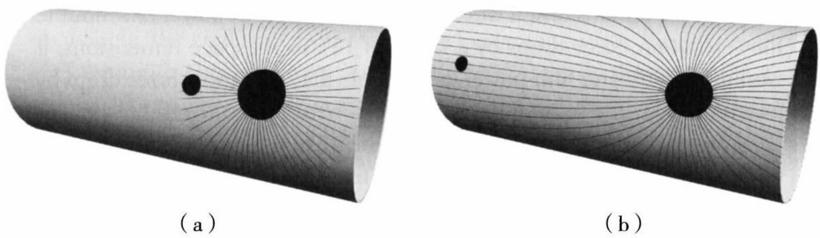



图13.6　（a）：当两个物体靠得很近时，引力表现的像是在二维空间中。（b）：当两个物体离得很远时，引力表现的才像是在一维空间中——此时引力为常数。

因而，如果你是这个宇宙中的一名实验物理学家，你发展了一套复杂精妙又准确的方法来测量引力，那这就是你的发现。当两个物体极其接近的时候，甚至比蜷曲维度的尺寸还要小的时候，两个物体间的引力就会随着两者间距的变大而减小，就如在一个具有两维空间的宇宙中应有的那样。但是，当两个物体间的距离大于额外的蜷曲维度的尺寸时，事情就不一样了。大于蜷曲维度尺寸时，不论两个物体间的距离如何变化，引力场线都不会有任何延展。一旦引力场进入第二个蜷曲维度，它就四散开来——它会在这个维度中达到饱和状态——因而在这样的距离上看引力场不会减小，如图13.6（b）所示。你可以将这种饱和与老房子中的管道相对比。如果有人在你正要洗净头上的洗发香波时打开了厨房水池的水龙头，你就会感觉水压骤降，因为两个水龙头都要流水。要是再有人同时打开洗衣间的水龙头，你就会觉得水压再次下降，因为有更多个水龙头同时分享水流。而一旦房子中的所有水龙头都打开了，水压就会保持不变了。尽管这时再也没法流出你所期待的令人能够放松的高压水流，可一旦水流在所有“额外的”水龙头间完全分流则淋浴头的水压就不会再发生变化了。类似的，一旦引力场在额外的蜷曲维度中完全延展开来，引力就不会随着距离的改变而改变。

从你的数据中你可以推断出两件事。首先，从两个距离很近的物体间的引力场反比于距离这一事实中，你会认识到宇宙中有两个空间维度，而不是一个。其次，从引力场为常数这一点——几百年积累下

来的实验数据——你可以得出结论，其中一维蜷曲了起来，蜷曲尺度大约就是引力场刚好能保持不变的距离。有了这样的结论，你就推翻了数百年乃至数千年来人们对于如此基本的空间维数的信念，而这看起来无疑比任何其他问题都更为惊人。

尽管出于视觉考虑，我们只是在低维的情况下讨论了这个问题，但对我们的三维宇宙来说，情况也是完全一样的。几百年来的实验数据告诉我们引力所遵循的乃是平方反比率，而这是空间有3个维度的有力证据。但1998年的时候，人们还从来没在小于一个毫米的尺度上做过检验引力强度的实验（今天，如我们先前提过的那样，对引力的检验已经到了比毫米还小一个量级的水平上）。这样的实验状况使斯坦福大学的萨瓦斯·蒂莫普洛斯，现在哈佛大学的尼玛·阿卡尼哈迈德，以及纽约大学的吉亚·杜瓦利共同提出，在膜世界方案中，额外维的尺寸可以大到毫米量级，并且还没有被实验探测到。这一激进的设想促使很多实验组开始着手研究亚毫米尺度上的引力，以期发现对引力平方反比率破坏的实验证据。到目前为止，直至1/10毫米量级上，尚未发现这样的结果。因而，即便依靠当前最为精良的引力实验设备，我们也没法证实我们是否真的生活在3-膜中，以及额外维度的尺寸是否大到了1/10毫米量级。

这样的现实是过去10年间最令人吃惊的事情。利用除引力之外的其他3种力，我们可以探索直至十亿分之十亿分之一米的尺度上的物理，可没人发现额外维度存在的任何证据。但在膜世界方案中，除引力之外的这3种力在探索额外维度方面将毫无用处，因为它们都被局限在3-膜之上。只有通过引力我们才能对额外维度有所了解，到今天为止，额外维度还是有可能厚到人们的头发丝那种程度，可我们却依然没办法在最为精良的设备上看到它们的存在。现在，就在你旁边，就在我旁边，就在我们每个人的旁边，就有可能存在着另一个空间维度——一个既不是左右，也不是前后，还不是上下的维度；一个虽然蜷曲起来但仍有可能大到吞下这张纸这么厚的东西的维度——而这个

空间维度仍在我们摸不着的地方。[71]

大额外维度与大的弦

膜世界方案将4种力中的3种限制在膜上，从而极大的放松了实验对额外维度尺寸的限制，但额外维还不是这个理论中唯一可以比较大的东西。通过吸取威滕的独到见解，乔·莱肯、康斯坦丁·巴查斯连同其他人、伊格纳缇奥斯·安东尼亚迪斯，以及阿卡尼哈迈德、蒂莫普洛斯、杜瓦利等人认识到额外维还有更为令人振奋的一面，低能的弦可能比先前认为的要大得多。事实上，这两个尺度——额外维的尺度与弦的尺度——是密切相关的。

我们在前面的章节中讲过，弦的基本大小是通过令其引力子振动模式按照观测到的强度传递引力而确定下来的。引力的微弱性使得弦非常的短，约为普朗克长度（$10^{-33}$厘米）。但是这一结论高度依赖于额外维的形状。而其原因则在于，在弦论或M理论中，我们在这个广大的三维世界中观测到的引力强度实际上由两个因素的相互影响导致的。一个因素是引力固有的基本强度，另一个因素就是额外维度的大小。额外维越大，进入额外维的引力就越多，从而使得出现在熟悉的维度中的引力就越弱。这就好比水管越粗则水压越小，因为在粗的水管中水可流动的空间更大。而更大的维度则会导致更小的引力，因为额外维度越大则引力延展的空间将会越大。

在确定弦的长度的最初计算中，做了额外维度小到普朗克长度的假定，这就使得引力几乎完全不能渗入额外维中去。在这样的假定下，引力之所以看起来很弱是因为它们本来就弱。但是现在，如果我们在膜世界方案中讨论问题并且允许额外维比之前认为的大很多的话，则人们观测到的引力的微弱性就不再意味着引力本来就弱。相反的，引力也可能是一种很强的力，而它之所以看起来很弱只是因为额外维度比较大，就像粗大的水管稀释了水压，大的额外维度也稀释了引力。顺着这种思路，如果引力真的比人们之前以为的大很多的话，

弦也应该比人们之前以为的大很多。

到今天为止，对于弦究竟可以有多长这个问题还没有一个独一无二的确定答案。弦的大小和额外维度的大小都可以在一个比之前认为的要大很多的范围内变化，有了这种新发现的自由后，弦的长度有很多种可能性。蒂莫普洛斯及其同事认为，现有的实验结果，不管是来自于粒子物理还是来自于天体物理，全都表明未激发的弦不可能大于十亿分之十亿分之一米（$10^{-18}$米）。而这个长度虽然比我们日常生活中的小还要小很多，但却已经比普朗克长度大10亿亿倍——差不多比之前认为的大10亿亿倍。如我们将要看到的那样，这就已经大到了可用下一代粒子加速器探测弦的信号的程度。

弦论遭遇实验

我们生活在一张大的3-膜中的这种可能性，当然，只是可能性。而且，在膜世界方案中，额外维度可能比之前认为的要大得多这种可能性——以及相关的弦也可能比之前认为的大得多的这种可能性——也只是可能性。但它们确实是令人异常鼓舞的可能性。当然了，即便膜世界方案真的正确，额外维的尺度以及弦的大小还是有可能在普朗克长度的量级。但是，在弦论或M理论中弦和额外维度的尺寸可以很大的这种可能性——刚刚超越今日的技术能力——真可算是异想天开。因为这意味着我们至少有机会在未来的几年中，看到弦论或M理论与可观测的物理联系起来从而成为一门实验科学。

这种机会有多大？我不知道，也没其他人知道。我的直觉告诉我这很可能没法成为现实，但我的直觉是基于15年传统的普朗克大小的弦与普朗克大小的额外维的框架下的工作中产生的。或许我的观念早就过时了。令人欣慰的是，这些问题的解决一丝一毫都并不取决于任何人的直觉。如果弦真的很大，或者说如果额外维真的很大，那么即将到来的实验将为我们带来激动人心的结果。

在下面的几章中，我们将讨论各种各样能够检验相对较大的弦和额外维度之可能性的实验，所以让我现在就开始吊起你的胃口。如果弦是十亿分之十亿分之一米大（$10^{-18}$米），则对应于图12.4中较高振动模式的粒子将不会有标准弦论中那样超过普朗克质量的巨大质量。相反的，它们的质量只会是一个质子质量的一千乃至数千倍，而这样的能量已经低到了目前正在CERN建设中的大型强子对撞机的能量范围。如果这些弦的振动通过高能碰撞而被激发，那么加速器上的探测器将像新年夜的时代广场上的水晶球一样亮起来。大量从未被看到过的新粒子将被产生出来，它们的质量彼此关联，就像大提琴的不同和音之间彼此关联那样。大量数据中的弦论信号将如亲笔签名般清晰。即使粗心的研究者也不会注意不到它们的存在。

而且，在膜世界方案中，高能碰撞甚至可能会产生出——记着这点——微观黑洞。尽管我们一般将黑洞视作外太空中的某种巨型结构，但即便早在广义相对论创立之初，人们即已认识到只要你将足够多的物质握在你的手中，你就有创造出一个小黑洞的可能性。而其所以不能成为现实则是因为没有什么人——也没有什么机器——有如此大的挤压力以至于能压出一个小黑洞。相反的，唯一可行的黑洞产生方式是这样的：一个质量巨大的星体的引力超过了星体核聚变过程所释放出来的向外的压力，使得星体向自身塌缩。但如果小尺度上引力的固有强度比之前认为的要大很多，那么只需要比之前认为的小很多的压缩力即可以产生出一个小黑洞。计算表明，大型强子对撞机通过高能质子对撞过程将正好获得足够的挤压力来制造大量的微观黑洞。[7]想想看吧，这有多么不可思议。大型强子对撞机可能会成为一个生产黑洞的工厂！这些黑洞如此之小，而且它们的存在时间也将极为短暂，不会给我们带来哪怕一丝一毫的威胁（很多年以前，史蒂芬·霍金证明所有的黑洞都将通过量子过程土崩瓦解——只不过大的黑洞瓦解得很慢，而小的黑洞则将瓦解得很快）。但是这些微观黑洞的产生，将为一些人类所遇到过的最稀奇古怪的思想带来坚实的证据。

膜世界宇宙学

当前研究的一个主要目的，世界范围内的科学家们（包括我在内）不断的热切追求的一个目标就是要用弦论或M理论的新观念来构建一个宇宙学理论。原因很明显：既不仅仅因为宇宙学要对付令其寝食难安的问题，也并不仅仅因为我们已然认识到自己熟悉的经验——比如时间之箭——的很多方面与宇宙诞生之初的状况紧紧联系在一起，更是因为宇宙学能为理论学家带来纽约为西纳特拉[72]提供的那种东西：证明自己的最好舞台。如果一个理论能够在宇宙最初时刻那种极端条件下取得成功，那它将在任何其他地方取得成功。

目前为止，根据弦论或M理论构建宇宙学还在进行之中，研究者们一般朝着两个方向进发。第一个方法更为传统，正如暴涨为标准的大爆炸理论提供了一个简洁但却意义深远的前端，弦论或M理论为暴涨提供了一个有关更早时期的意义可能更加深邃的前端。在这个图景中，代表我们对宇宙最初时刻的无知的那一片模糊将因为弦论或M理论而变得清楚起来，在那之后，宇宙大戏将按照在前面的章节中讲过的取得了极大成功的暴涨理论设计的脚本一幕幕上演。

尽管在这一图景所要求的某些细节方面已经取得了一些进展（比如对为什么宇宙的空间维度中只有3个会经历膨胀的理解，以及某些数学工具的开发——这些数学工具可以用来分析暴涨之前无空或无时的领域），但成功一刻尚未来到。直觉上，尽管在暴涨宇宙学看来，越早时刻的可观测的宇宙会变得越小——因而也会变得更热、更密、更具活性——弦论或M理论却通过引入一个最小尺寸（如我们在上一章中的“小尺度上的宇宙结构”那一节中讲过的那样）来驯服有着蛮横狂暴的行为方式（用物理学的语言来说是，具有“奇异性”）的早期宇宙，而在我们所引入的这个小尺寸下，起作用的将只是一些新的，不那么具有奇异性的物理量。这种思考正是弦论或M理论在合并广义相对论与量子力学方面取得成功的关键所在。而且，我感觉我们很快就可以决定如何将这种思考应用于宇宙学范畴。但是，到目前为止，那

模糊的一片还是一如既往的模糊，没人知道明晰什么时候才会来到。

第二种方法采用了膜世界方案，在其最为激进的实现中，抛出了全新的宇宙学框架。我们还远不知道这种方法能不能克服数学细节上的困难，但作为一个例子，它很好地向我们展示了基本理论上的突破如何在常见领域上留下新的足迹。这一新思想被称为循环模型。

循环宇宙学

从时间的角度看，我们感受的普通体验通常只有两种现象：一种有清楚的开始，中间，以及结尾（这本书，一场棒球赛，人的一生）；还有一种循环往复周而复始（四季的变化，日升日落，拉里·金的婚礼[73]）。当然，仔细推敲的话循环现象也有开始与结尾，因为循环一般也不是永恒存在。每天日升日落——地球一边自转一边绕着太阳公转——的过程只有50多亿年。在那之前，太阳和太阳系还没形成。而且总有那么一天，比方说50亿年后或者多少年后，太阳会变成红巨星，吞噬掉太阳系内的一切星体，地球也不能例外，那时就再也没有日升日落了，至少在太阳系没有了。

但这些都是现代的科学认识。对古人来说，循环现象看起来像是永恒不变的。而且对于很多人来说，循环现象运转着他们的生活，一次又一次的重新开始，实在是再正常不过的事情。每天每季的周而复始为人们设定工作与生活的节奏，所以无怪于一些最为古老的宇宙学会将世界的演变过程想象成循环过程。循环宇宙模型从拥有一个开始，中间，以及结尾的过程中解脱出来，视世界随时间变化如月亮随月相改变一般：在完成一个完整序列之后，时机成熟，世间万物重新开始另一次循环。

自从广义相对论被发现以来，人们已经提出了大量的循环宇宙模型，其中最为著名的是加州理工学院的理查德·托尔曼于20世纪30年代开发的版本。托尔曼提出，人们观测到的宇宙膨胀可能会慢慢减

缓，最终停下来，之后将经历一个宇宙慢慢变小的收缩期，但是最后不会终止于自身的猛烈聚爆。托尔曼认为，宇宙将会反弹：空间会缩到某一极小的尺度上，然后反弹，再开始新一轮的膨胀，之后再次收缩，如此循环下去。宇宙会永不停歇地重复着这种循环过程——膨胀，收缩，反弹，再膨胀——这将绕开令人头疼的起源问题：在这样的方案中，起源的概念是没有意义的，因为宇宙一直就是那个样子并将永远那样下去。

但是托尔曼认识到，追溯回去，循环可能重复了几次，但不确定。原因在于，在每一次循环过程中，根据热力学第二定律，平均说来熵都只能增加。[8]而根据广义相对论，每一次循环开始时熵的总量将决定这次循环会持续多久。更多的熵意味着向外膨胀过程慢慢停下来并转成向内收缩之前的膨胀周期更长一些。因而每一次后续循环过程都会比前一次更久些；这等于说，前面的循环周期应该越来越短。用数学分析一下便会发现，循环过程周期的连续变短意味着这种循环不能推演到无穷远的过去。所以，即使在托尔曼的循环理论框架下，宇宙也会有一个开端。

托尔曼的想法带来了球形宇宙模型，但是正如我们所见，它早就被实验观测排除掉了。但是循环宇宙模型的一种新的实现，与平直宇宙有关，近年来在弦论或M理论的框架下发展起来。这一想法来自于保罗·斯坦哈特及其在剑桥大学的同事尼尔·塔洛克（这一想法的提出离不开他们先前与柏特·欧弗拉特、内森·塞博格以及贾斯汀·霍里的合作成果）。他们提出了一种新的驱动宇宙演化的机制。[9]简单说来，他们认为，我们生活在3-膜中；而我们的这个3-膜与邻近的平行3-膜每隔几万亿年就会发生一次剧烈碰撞。而来自于碰撞的“爆炸”开启每次新的宇宙循环。

这一想法的基本构建如图13.7所示，很多年前由哈罗瓦和威滕在非宇宙学的框架下提出。哈罗瓦和威滕当时正在试图完成威滕提出的5种弦论的统一，他们发现如果M理论中的7个额外维度中的一个有某

种非常简单的形状——不是图12.7中的圆环，而是图13.7中那样很小的线段——并且被所谓的世界尽头之膜夹在中间，就像一本书被两片书立夹在中间那样，那么就可以在杂化E型弦论与其他弦论之间建立直接联系。他们如何找到这种联系的细节在这里既不明显也不重要（如果你感兴趣，可以参考《宇宙的琴弦》第12章）；这里关键的是，有一个起点自然地从理论本身中冒了出来。斯坦哈特和塔洛克就利用这些发现提出了他们自己的宇宙理论。

特别是，斯坦哈特和塔洛克将图13.7中的每张膜想象成有3个空间维度，而两张膜之间的线段就是第4个空间维度。剩下的6个空间维度蜷曲为卡拉比丘空间（未在图中画出），而这个卡拉比丘空间需要有恰当的形状以便用弦的振动模式解释已知的粒子种类。[10]我们直接感知到的宇宙对应于两张3-膜之一；如果你愿意，你可以将第二张3-膜想象成另一个宇宙，那个宇宙中的居民，如果存在并且假定其实验技术与知识水平并未远远超越我们的话，也只知道空间有3个维度。在这种设定下，另一张3-膜——另一个宇宙——其实就在旁边。它就盘旋在离我们不足1毫米远的地方（这里的间隔指的是第4个空间维度中的距离，如图13.7所示），但是由于我们的3-膜如此之厚而且我们感受到的引力如此之弱，所以我们没有另一张3-膜存在的直接证据，而那个宇宙中的居民也不会有我们存在的直接证据。

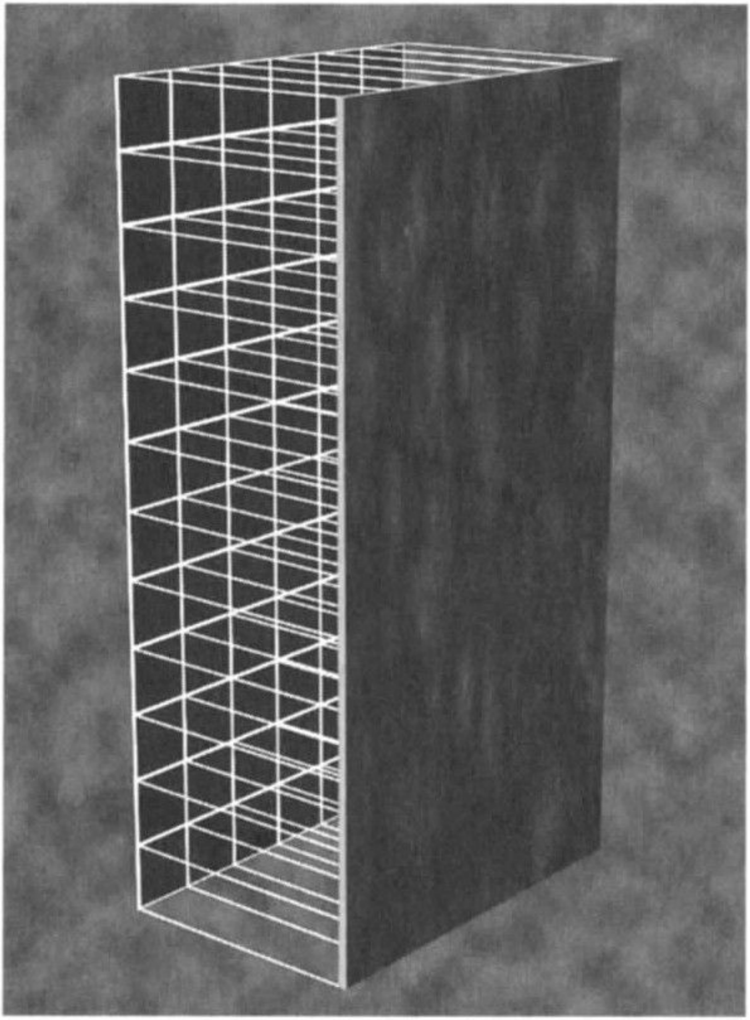


图13.7 间距很短的两片3-膜。

但是，根据斯坦哈特和塔洛克的循环宇宙模型，图13.7画的并不是它一直以来以及未来可能的样子。与之相反，在他们的方法中，两个3-膜彼此吸引——就像被细细的橡皮带连起来了似的——而这就意味着每一个3-膜都在驱动着另一个3-膜的宇宙演化：这些3-膜注定将处于无限循环的碰撞，反弹，以及再次碰撞，一而再再而三乃至永远

的生成它们各自膨胀的三维世界。可以看看图13.8来弄明白究竟是怎么回事，这张图一步步地展示了一次完整的循环。

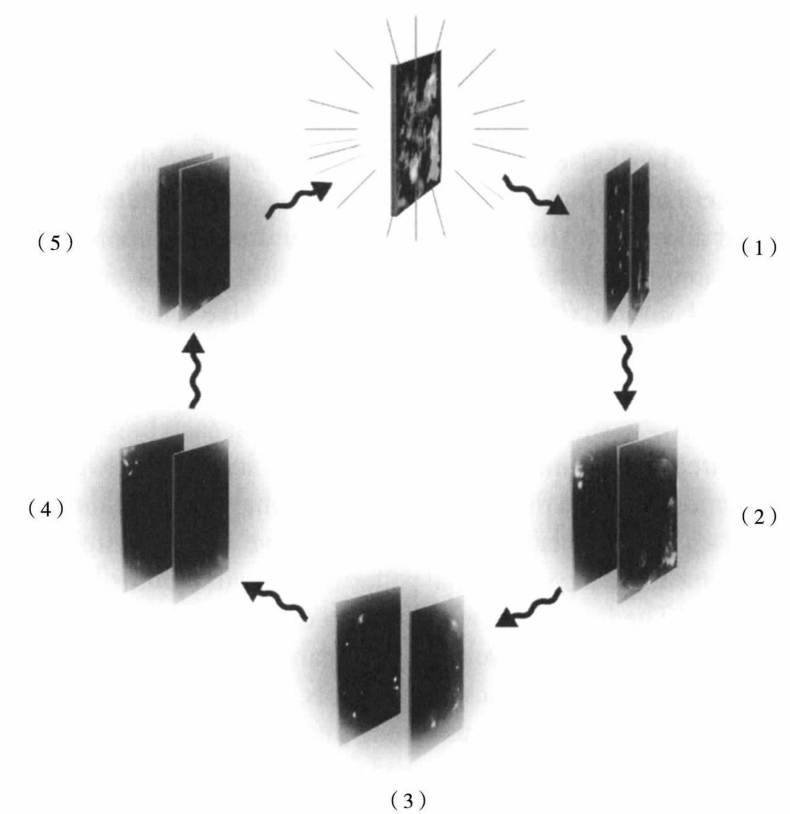



图13.8　循环宇宙模型中的不同阶段。

第一阶段，两个3-膜刚刚撞在一起，正要弹开。碰撞产生的巨大能量将巨量的高温辐射以及物质沉积在了每一张弹开的3-膜中，而斯坦哈特和塔洛克认为——这一点非常重要——这些物质与辐射的具体性质同暴涨模型中产生出来的物质与辐射的性质有着极为类似的特征。尽管在这点上还存在着某些争议，斯坦哈特和塔洛克依然可以宣称由两张3-膜的碰撞导致的物理条件极其类似于我们在第10章中讨论过的传统方法中的暴涨膨胀爆发之后的瞬间的物理条件。所以毫不奇

怪，对于我们的3-膜上的假想观测者来说，循环宇宙模型中的下几个阶段在本质上将与图9.2（现在可以将这张图解释为描述了某张3-膜中的宇宙演化）所示的标准方法中的相应阶段完全一样。也就是说，当我们的3-膜从碰撞中弹出的时候，它就开始膨胀并冷却，而原初等离子体逐渐汇聚形成恒星与星系这样的宇宙结构，如图中第二阶段所示。然后，受我们在第10章中讨论过的近年来对超新星的观测启发，斯坦哈特和塔洛克将他们的模型重新设定，以便在循环进行到70亿年的时候——第3阶段——普通物质中的能量以及辐射会因膜的膨胀而得以足够稀释，从而使暗能量成分赢得足够优势，并通过其负压，驱动宇宙进入加速膨胀时代（这就要求对一些细节任意调节，但是却会使模型与观测相符合，而在循环宇宙模型的支持者看来，这种符合正是提出这个模型的动机）。此后再过70多亿年，我们人类出现在地球上——至少按现在这个循环看是这样——感受着早期的加速阶段。然后在接下来的差不多3万亿年间，我们的3-膜始终在膨胀。在这样漫长的岁月中，我们的三维空间被极大地拉伸，物质与辐射被充分地稀释，这使得膜世界看起来空空荡荡又天下大同：这就是第4阶段。

到那时，我们的3-膜就完成了初始碰撞后的反弹，再次朝着另一张3-膜飞去。当我们离另一次碰撞越来越近的时候，我们的膜上的弦的量子涨落将使整个空荡荡的宇宙泛起细小的波纹，这是第5阶段。我们的3-膜继续加速，小小的量子波纹变得剧烈起来；猛然间，毁灭性的碰撞来到了，我们撞向另一张3-膜，碰撞之后我们又弹开，另一次循环开始了。量子涨落将记录下碰撞产生的辐射与物质的各向异性，就像在暴涨理论中那样，这些对完美的各向同性的偏离逐渐聚集起来最终形成恒星与星系。

这就是循环宇宙模型（也被称为大碰撞模型，big sPlat）中的几个主要阶段。其基本假定——膜世界的碰撞——与已经取得成功的暴涨理论完全不同，但是却在几个重要方面与暴涨理论是相通的。这两个理论都要依靠量子扰动来生成最初的不均匀性，这是两者间非常重

要的类似之处。事实上，根据斯坦哈特和塔洛克的论证，掌控循环宇宙模型中的量子涨落的方程与暴涨理论中的极其类似，因而两个理论所预言的不均匀性也几乎完全一致。[11]而且，尽管循环宇宙模型中没有暴涨，但却有一个长达万亿年温和加速膨胀时期。两者间真正的区别就在于一个急躁一个耐心；暴涨理论在瞬间完成的事情，循环宇宙模型花了万亿年才做完。既然循环宇宙模型中的碰撞并不是宇宙的起源，那就有可能在上一次循环后3万亿年间慢慢消融宇宙学问题（比如平坦性疑难和视界问题）。每一个循环后期那无数年和缓而稳定的加速膨胀将我们的3-膜拉伸得既干净又平整，而且，除了微小却重要的量子涨落，整个宇宙空间均匀一致。因而，每一次循环那漫长的最后一个阶段——紧随其后的就是下次循环开端的大碰撞——产生的宇宙环境看起来竟与暴涨理论中的瞬间猛烈膨胀所产生的宇宙环境如此类似。

简评

在其发展的现阶段，暴涨与循环模型为我们带来了富于启发性的宇宙学框架，但是两者中的哪一个都不能给我们一个完整的理论。对宇宙最初时刻的主要条件的无知，迫使暴涨宇宙学的支持者们只能不经理论判定的去假设暴涨所需要的初始条件。如果真的有那些初始条件，那么暴涨理论就会解决大量的宇宙学难题，并且启动时间之箭。但这些成功都要以暴涨发生为先决条件。而且，暴涨宇宙学没法天衣无缝地嵌入弦论中，也不可能同时与量子力学与广义相对论保持一致。

循环宇宙模型也自有其短处。如同托尔曼模型，考虑到熵的问题（以及量子力学[12]），循环宇宙模型的循环不会永远持续下去。相反的，循环开始于过去的某个特定时间，所以，就像暴涨模型遇到的问题一样，我们也需要解释第一次循环是怎么开始的。如果我们解释了循环宇宙模型的开端，那么这个理论，如同暴涨理论，也将解决关键的宇宙学问题，并使时间之箭从每一次低熵的开天辟地，连续的经历

图13.8所示的那些阶段。但是，按我们现在的认识，循环宇宙模型还没法解释宇宙怎样以及为什么使自己满足图13.8所需的必要条件。比方说，为什么会有6个维度将自己蜷曲成特定的卡拉比丘形，而同时还有一个维度忠实的保持着自己作为两个膜的间隔的形状？两个作为世界尽头的3-膜为何会完美的联系起来，又为什么会以恰好的力吸引彼此以至于我们在图13.8中所描述的过程可以进行下去？而且，尤为重要的是，当两张3-膜按循环宇宙模型版的大爆炸撞到一起时，到底发生了什么？

关于最后一个问题我们得说，比起暴涨宇宙学在时间零点遇到的奇异性，循环宇宙模型中的开天辟地问题要少得多。不同于暴涨宇宙学中所有的空间维度都被无限压缩，在循环宇宙模型中，只有一个维度被极度压缩；膜本身在每个循环过程中处于整体扩张之中，而不是压缩。斯坦哈特、塔洛克及其合作者们认为，这种情况意味着膜本身具有有限温度以及有限密度。但这只是一个具有高度不确定性的结论，因为到目前为止，还没有人能够得到更好的方程，并指出两个膜撞在一起时究竟发生了什么。事实上，目前为止的分析表明，循环宇宙模型中的开天辟地也遭受着暴涨宇宙学在时间零点遇到的问题：数学工具破产。因而，宇宙学的奇异开端——它到底是宇宙的开端呢，还是我们目前这次循环的开端呢——还是需要一个严格的解决方案。

循环宇宙模型最吸引人的性质，是它将暗能量以及观测到的加速膨胀纳入自己的体系中的方式。1998年，人们发现宇宙正在加速膨胀时，这令大多数宇宙学家和天文学家困惑不解。尽管只要假定了宇宙中有合适数量的暗能量，暴涨宇宙学就能解释这种加速膨胀，但加速膨胀看起来总像是粗笨的附属品。与之相比，在循环宇宙模型中，暗能量的角色就自然和重要得多。3万亿年的缓慢而稳定的加速膨胀，对清除各种麻烦，稀释可观测宇宙使之近乎一无所有，重新设置所有条件准备下次循环十分关键。从这种角度看，暴涨宇宙学和循环宇宙模型都依赖于加速膨胀——暴涨模型接近其开端而循环宇宙模型中靠

近每次循环的尾声——但只有后者有直接的观测支持（记住，循环宇宙模型的设计就是要使我们刚刚进入3万亿年的加速膨胀阶段，而这样的加速膨胀是最近才被观测到的）。加速膨胀是循环宇宙模型的标志，但是，这也就意味着，假如未来的实验观测又表明不存在加速膨胀，那暴涨宇宙学还是能存活下来（尽管宇宙能量预算中那丢失的70%又要以新面孔出现了），可循环宇宙模型就不能了。

时空的新图景

膜世界方案与脱胎于其中的循环宇宙模型都是高度理论性的想法。我之所以在这里对其进行讨论并不是因为我认为它们正确，而是因为，我想展示一下按照弦论或M理论所启发的新方式，如何思考我们生活于其中的空间及其演化。如果我们生活在3-膜中，那么几个世纪以来有关三维空间的形体存在的老问题就会得到确定的答案：空间是一张膜，因而空间真的是某种实在的东西。而膜并没什么特别之处，因为在弦论或M理论的高维空间中漂浮着其他很多各种维数的膜。如果我们的3-膜的宇宙演化就是与邻近3-膜的重复碰撞，那么我们所知道的时间只能跨越宇宙众多循环——一次大爆炸，紧接着另一次，接着又一次，循环下去——中的一次。

对我来说，它是一个既激动人心又令人谦卑的版本。空间和时间或许远超我们的预期；如果真是那样，则我们所认为的“万物”可能只不过是更为丰富的实体的小小组成部分。

# 第5部分
# 真实与想象

## 第14章　上天入地

### 关于空间与时间的实验

从阿格里琴托的恩培多克勒用土、空气、火和水解释宇宙到今天，人类对空间和时间的理解走过了漫长的旅程。我们所取得的很多成就，从牛顿理论到20世纪的革命性发现，都由于理论预言与实验结果的精确符合而得以验证。但时间推移到了20世纪80年代中期，我们似乎成了过去辉煌的受害者。在科学家们孜孜不倦的好奇心的驱动下，当代理论已进入实验技术无法触及的领域。

不过，实验学家们当然不会就此甘心，靠着勤奋和运气，他们找到了一些检验当代最前沿思想的方法，这些方法将在未来的几十年间付诸实践。我们在本章中将会看到，一些已经启动或正在计划中的实验将帮助我们弄清额外维度存在与否，暗物质和暗能量的组成，质量的起源与希格斯海，早期宇宙学的某些方面，超对称的相关内容，甚至弦论的真实性。所以，要是我们再有一点运气的话，一些在统一理论、空间与时间的性质以及宇宙的起源等方面富于想象力和革命性的思想将最终得以检验。

### 陷入困境的爱因斯坦

在为建立广义相对论而艰苦奋斗的那10年间，爱因斯坦从各种源头寻求灵感。其中，由18世纪的著名数学家卡尔·弗雷德里希·高斯、

詹诺斯·波尔约、尼古拉·罗巴切夫斯基和格奥尔格·伯恩哈德·黎曼等人所创立的关于弯曲形状的数学带来的影响最为深远。我们在第3章曾经讨论过，欧内斯特·马赫的思想也曾为爱因斯坦带来过灵感。还记得马赫所提出的空间的关系概念吗？对于马赫来说，空间只是一种指定不同物体彼此之间的相对位置的语言，其本身并不是一种独立实体。起初，爱因斯坦是马赫观点的热情拥护者；因为在当时看来，马赫的观点最具相对论性。但是随着对广义相对论理解的加深，爱因斯坦认识到广义相对论与马赫观点并不能完全相容。根据广义相对论，在牛顿的那个在真空中旋转的桶中，水面会成凹陷状；这一点与马赫观点相矛盾，因为水面凹陷相当于暗示着绝对的加速概念。不过即使这样，广义相对论还是在很多方面同马赫的观点相一致，在未来的几年间，酝酿了差不多40年、造价高达5亿美元的实验将检验马赫原理中最著名的一个性质。

将在这个大型实验中研究的物理可以追溯到1918年。那一年，奥地利物理学家约瑟夫·兰斯和汉斯·塞林利用广义相对论证明：就像有质量的物体会使空间和时间弯曲——想想蹦床上的保龄球，旋转的物体也将拖曳其周围的空间（与时间）——这次你可以想想掉进果酱桶中的旋转石块。这一现象被称之为框架曳引；我们来举个例子以说明这一现象。想象一个向着急速旋转的中子星或者黑洞自由下落的小行星，它会被卷入旋转空间的漩涡中，在其下落的过程中会被拖动着旋转，这种效应就是框架曳引。而这种效应之所以被称之为框架曳引是因为从小行星的角度看——从其参考系看来——它并没有被拖动旋转。不但没有旋转，小行星甚至是沿着空间格子按直线下落；但是由于空间形成了漩涡（如图14.1所示），格子变得扭曲。所以“直线下落”的概念和你以往在平直空间中形成的印象有所不同。

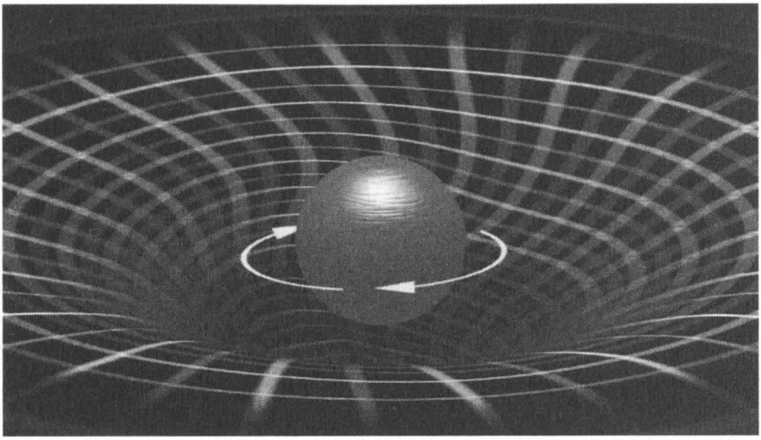


图14.1　有质量的旋转物体会拖曳其周围的空间——可以随便放入任何东西的框架。

为了看清楚框架曳引效应和马赫原理之间的联系，让我们来试想一种由巨大的有质量的旋转空心球引起的框架曳引效应。1912年爱因斯坦（甚至在其完成广义相对论之前）首先进行了有关计算，1965年戴尔特·布里尔和杰弗里·科恩对爱因斯坦的讨论做了重要扩充，最后，1985年，德国物理学家赫伯特·菲斯特和K·布劳恩彻底完成了这一计算。这些物理学家的相关工作表明，空心球内部的空间会被旋转运动拖曳，形成漩涡状的旋动。[1]如果被固定住的桶内装满了水——所谓的“固定住”指的是从远处的参考点看——并被放到这样的一个旋转空心球内，那么根据计算，旋转的空间会对处于静态的水施加力的作用，使水相对于桶旋转起来，水面凹陷下去。

这样的结果肯定会令马赫非常高兴。尽管他可能会不喜欢“旋转的”空间这样的说法——因为这样的术语将时空视作某种东西——但他肯定会对空间与桶之间的相对旋转运动导致水面形状改变非常满意。事实上，如果外壳具有足够多的质量，多到足以与整个宇宙的质量不相上下，那么根据计算可知，无论你将这一过程看做空心球绕着桶旋转，还是看做桶在空心球内旋转，都没有关系。正如马赫所主张的那样，唯一有关系的是两者之间的相对运动。我在上文中提到的这一计算没有用到除广义相对论之外的任何东西，所以它可算作是爱因

斯坦理论中一个具有明显的马赫性质的例子。（不过，根据标准的马赫式推理，在无限大的空宇宙中旋转的桶里，水面会保持平面不变；但是根据广义相对论所得出的结论则并非如此。菲斯特和布劳恩的计算告诉我们，质量足够大的旋转球面能够完全隔绝通常情况下球面外的空间所带来的影响。）

1960年，斯坦福大学德莱奥纳德·席夫和美国国防部的乔治·普夫分别独立提出，框架曳引的广义相对论预言可以利用地球的自转实验检验。席夫和普夫认识到，在牛顿理论中，悬浮在高于地球表面的轨道上的回旋陀螺仪——连在一根轴上的旋转轮——会一直指向固定的方向。但是根据广义相对论，回旋陀螺仪的轴会由于地球的空间曳引而很轻微的旋转。与菲斯特和布劳恩的计算中用到的假想空心球相比，地球的质量非常之小，相应的，地球的旋转导致的框架曳引效应也非常之小。详细的计算表明，如果回旋陀螺仪的转动轴初始指向选定的参考星，一年之后，缓慢的空间旋转会使回旋陀螺仪的指向改变十万分之一度。这一度数大约是钟表上的秒针在两百万分之一秒中转过的角度，所以，这样的探测无疑是对当代科学技术及工程能力的巨大挑战。

经过40年的发展，产生了近百篇博士论文之后，由弗朗西斯·艾弗里特领导、NASA资助的斯坦福组已经准备启动这一实验，在未来的几年间，漂浮在400千米之外的太空中、装备着有史以来最稳定的回旋陀螺仪的引力探测器B卫星将开始探测由地球的自转导致的框架曳引效应。一旦这一实验取得成功，它将成为有史以来最精确的广义相对论实验，并为马赫效应带来第一个直接证据。[2]该实验也可能探测到与广义相对论的预言有所偏离的结果，这种可能性也同样令人兴奋。如果真的得到这样的结果的话，广义相对论中这小小的不和谐之音将使我们初窥迄今未见的时空性质。

捕获波

广义相对论告诉我们的一件重要事情是质量和能量可以使时空结构发生蜷曲。我们在图3.10中曾展示过太阳周围的弯曲情况。但是，静态的图片不能说明所有的问题，它有一定局限性，因为静态图片不能告诉我们当质量和能量发生转移或者以某种方式改变其自身构成时，空间的蜷曲将如何演化。[3]要是你非常老实地站在一张弹簧床上，弹簧床就会保持固定的弯曲形状；而一旦你开始乱蹦乱跳，弹簧床就会跟着上下起伏。广义相对论也能给出与此类似的预言，如果物质处于完美的静止状态，空间就会保持固定的蜷曲形状，如图3.10所示的那样；但是一旦物质运动起来，空间的结构就会有所起伏。爱因斯坦在1916～1918年间认识到这一点，就在那个时期，他利用时新的广义相对论方程证明——就像在广播天线上来来回回的电荷会产生电磁场一样（无线电波与电视信号就是这样产生的）——物质的猛烈涌动（比如超新星爆发）会导致引力波的产生。因为引力就是曲率，所以引力波就是曲率波。将一块鹅卵石投入池塘会激起层层涟漪，向外扩散；旋转的物质会导致向外传播的空间涟漪。根据广义相对论，遥远天际中的超新星爆发就像投入时空这片巨大的池塘中的鹅卵石一样，会激起层层涟漪，如图14.2所示。这张图展示了引力波与众不同的一个重要性质：不同于电磁波、声波和水波这些穿行于空间传播的波，引力波就通过空间自身传播。引力波传播的就是空间本身的扭曲波动。

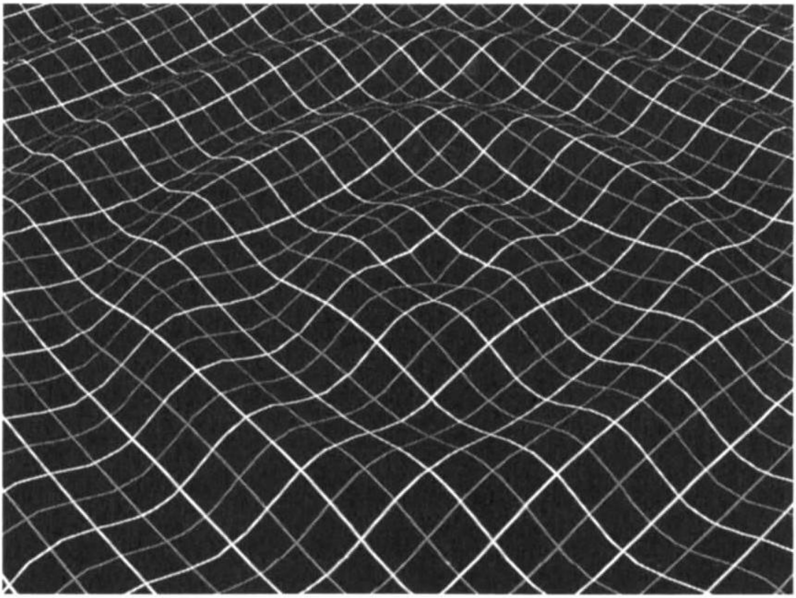


图14.2　引力波就是空间结构中的波纹。

尽管现在的人们已经把引力波当做是广义相对论的预言接受了下来，但是，有关这一课题的研究一直含混不清，饱受争议。造成这种情况的部分原因在于有些人死守着马赫哲学不放。如果广义相对论与马赫的思想完全协调一致，那么“空间的几何”就只能算是一种可以很方便地表述一个有质量的物体相对于其他物体的位置与运动的语言。以这种方式思考的话，真空这一概念将意味着真正的空空如野，那么讨论真空本身的波动还有什么合理性呢？很多物理学家曾试图证明假想中的空间中的波只不过是对广义相对论的数学的一种曲解。但是，所有的理论分析最后总是归结到正确的结论：引力波是真实的，空间可以波动。

引力波的波峰波谷川流不息，在某一方向上拉伸空间——及其中的一切，再在另一垂直方向上压缩空间——及其中的一切。如图14.3所示。原则上，你可以通过反复测量多个不同位置之间的距离，发现

这些距离之间的比率有所变化来探测到引力波。

但在实践中，没有人能够完成这样的任务，因而从未有人直接探测到引力波（即使这样，我们仍然有间接的有力证据支持引力波的存在[4]）。这一实验的困难之处在于，经过的引力波所带来的空间扭曲效应太小。1945年7月16日在新墨西哥州的Trinity[74]试爆的原子弹产生了相当于2万颗TNT炸药爆炸所产生的能量，所发出来的光异常强烈，以至于数英里之外的目击者仍需要带上护具以防止眼睛被原子弹产生出来的电磁波伤害。然而，即使你就站在放置原子弹的百英尺高铁塔之下，由爆炸所产生的引力波也仅仅会使你的身体拉伸不足原子直径的长度。引力波所带来的波动就是如此之弱，这无疑暗示着探测引力波是对技术能力的巨大挑战。（因为我们也可以将引力波看做是数目巨大的引力子按同样的方式运动——就像电磁场是由数目很多的光子组成——所以引力波的影响之弱也暗示着探测到单个引力子非常困难。）

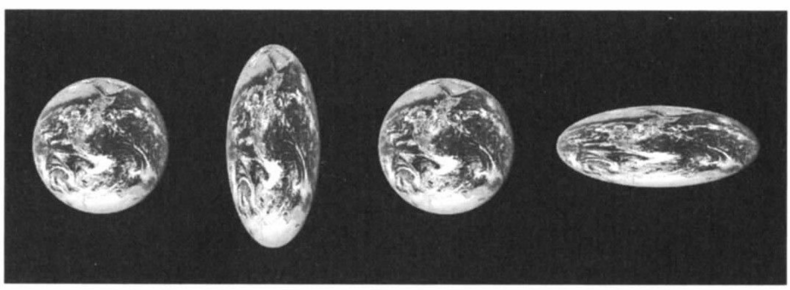


图14.3　引力波穿过物体的时候，会忽而这样忽而那样地拉伸物体（在这张图片中，典型引力波的扭曲尺度被极度的放大了）。

当然，我们感兴趣的并不是探测到原子弹爆炸所产生的引力波。但即便我们感兴趣的是能量要大得多的天体源产生的引力波，要探测到其存在也并非易事。天体源距离我们越近、质量越大，并且有关的运动能量越高、运动越猛烈，我们所接收到的引力波就将越强。但是，就算在10000光年远的距离上有一颗恒星变成了超新星，传到地

球上的引力波的强度也就只能使1米长的杆拉伸千万亿分之一厘米——大约只是原子核尺度的百分之一。所以，除非在距离我们相对近些的位置上发生了某种出人意料的超大规模天体物理事件，否则的话，我们就只能通过发展能够探测在难以置信的小尺度上的尺寸变化的实验装置才能探测到引力波。

设计并建造了激光干涉仪引力波探测器（LIGO）（由美国国家科学基金出资，加州理工学院和麻省理工学院联合运作）的科学家们接受了这一挑战。LIGO受人瞩目，具有令人难以相信的精度。它由两个空心管组成，每一个有4千米长，1米多宽，这两个管排成巨大的L形。激光在每一个管内的真空通道中同时照射，并被高度抛光的镜子反射，人们就用这样的装置高精度地测量相对长度。这一装置的设计思想在于，经过的引力波会使某一根管子相对于另一根有所拉伸，一旦这种拉伸足够大，科学家们就能探测到引力波的存在。

这样的管子之所以要造得很长是因为引力波带来的拉伸和压缩具有累加性。也就是说，如果引力波能把某个长4米的东西拉伸$10^{-20}$米，那么它就同时能把另一个4千米长的东西拉伸$10^{-17}$米。因而，探测的空间间隔越长，测得其长度发生改变就越容易。为了能够更好地利用这一点，LIGO实验实际上是让激光束在置于每根管子相反两端的镜子之间来来回回地反射上百次，这样可以利用每束激光实际探测大约800千米的长度。有了这样聪明的技巧和先进的工程技术，LIGO有能力探测到管中如人类头发丝的万亿分之一的长度——原子的亿分之一——上的改变。

对了，这样的L形装置实际上有两个。一个坐落于美国路易斯安那州的利文斯顿，另一个位于2000千米之外华盛顿州的汉福德。远方的天体物理喧嚣通过引力波使地球感受到的时候，会带给两个探测器相同的影响，我们在一个探测器上看到的引力波应该与在另一个探测器上看到的一样。用两台探测器进行这样的交叉检验非常有必要，因为即使人们采用了种种手段屏蔽探测器，那些生活中常见的振动（比

如卡车通过时的隆隆声，链锯的嗡嗡声，轰然倒下的大树，如此等等）还是有可能会冒充引力波。而要求相隔很远的两个探测器上得到相同的结果则会排除掉这些可能的错误信息。

对于包括超新星爆发、非球形中子星的旋转运动，以及两个黑洞之间的碰撞在内的一大类可能产生引力波的天体现象，研究人员们都仔细计算了其引力波的频率——每秒钟内通过探测器的波峰波谷数。没有这些信息的话，实验家们就是在大海捞针；有了这些信息的话，实验家们就可以将他们的探测器聚焦到物理上感兴趣的波段。严格上讲，计算表明某些引力波的频率在每秒几千次左右；这些波要是声波的话，那它们就在人类的听觉范围内。中子星听起来就像音调急速升高的叽喳声一样，而一对碰撞的黑洞听起来则像被风当胸猛吹的麻雀发出的颤音一样。振荡于空间结构中的引力波就像丛林中的杂音一样，如果一切按计划进行，LIGO将是第一件能够收听这些声音的器具。[5]

使这一切如此令人激动的原因在于，引力波最大程度地展现了引力的两个主要性质：弱与无处不在。在所有的4种力中，引力与物质的相互作用最为微弱。正是这一点使得引力波能够穿过光无法通过的物质，使得我们能够触及以前隐藏起来的天体物理领域。而且，因为万物都受引力掌控（其他的力则并非如此，比如电磁力就只对带电物体有作用），所以世间的一切都有可能产生引力波以及可观测的信号。在这种意义上，LIGO可算是人类探索宇宙的转折点。

曾几何时，人类只能大睁双眼，仰望星空。17世纪，汉斯·利伯希75和伽利略改变了一切；在望远镜的帮助下，宇宙的广阔景象进入了人类的视野。很快地，人类就认识到可见光只是整个电磁波段中很窄的一块。20世纪，在红外、无线电、X射线以及伽马射线望远镜的帮助下，宇宙在我们的眼前变得焕然一新，我们看到了用肉眼不能看到的波段处的宇宙景象。现在，21世纪到了，天空的疆域在我们的面前再一次扩大了。利用LIGO及其未来的升级版，76我们将能以一种全新

的方式重新审视宇宙。我们没有使用电磁波，而是使用了引力波；没有利用电磁力，而是利用了引力。

为了更好地体会这种新技术可能带来的革命性进展，我们可以想象有一群外星世界的科学家刚刚知道了如何探测电磁波——光，他们还在思考这一发现在短期内将如何改变他们对宇宙的认识。我们也正好处在第一次探测到引力波的前夜，与那些外星科学家所处的情况很类似。我们仰望这个宇宙已经几千年了，现在，人类有史以来第一次，我们得到了聆听它的机会。

寻找额外维度

1996年之前，在大部分将额外维度的想法纳入其中的理论模型中，额外维度的尺度都是普朗克量级的（即$10^{-33}$厘米）。这样的量级比当前实验可能触及的区域小了足足17个量级，如果技术上没有什么奇迹发生的话，普朗克尺度上的物理不可能进入我们的研究领域。但是如果额外维度很“大”，大于万亿亿（$10^{-20}$）分之一米——大约是原子核尺度的百万分之一，那么普朗克尺度的物理就有可能成为我们的研究对象。

正如我们在第13章中讨论过的那样，如果有一些额外维度“非常大”——大到几个微米的水平上——对引力强度的精确测量就将揭示它们的存在。这样的实验已经进行了几年，技术上也是日新月异。到目前为止，人们还没有发现偏离三维空间中平方反比率的迹象，研究人员正在进一步探索更小的尺度。一旦发现偏离的信号，物理学的基础将被猛烈的撼动。这样的信号会提供只对引力开放的额外维度存在的坚实证据，并对膜世界机制和弦论或M理论提供强劲的间接证据。

如果额外维度不小，但又并不是非常大，那么精确的引力实验就可能探测不到它们的存在，但是其他的间接方法还有可能起作用。比如说，我们在前面的讨论中曾经提到过，额外维度的存在暗示着引力

的内禀强度可能比之前认为的要大。引力在观测上的微弱性可能是由于引力部分渗透到了额外维度中导致的，而不是由于其本身微弱导致的；在很小的尺度上，引力还不能进入额外维度，引力可能很强。由此导致的其他推论姑且不提，单说产生小黑洞所需要的质量和能量，就有可能比之前在一个引力本身就很弱的宇宙中预计需要的能量少很多。在第13章中，我们曾经讨论了这样的微观黑洞在大型强子对撞机——现在正在瑞士的日内瓦建造中的粒子加速器，预计于2007年完工[77]——上的高能质子质子碰撞过程中产生的可能性。这样的前景激动人心。肯塔基大学的阿尔佛雷德·夏皮尔和加利福尼亚大学欧文分校的乔纳森·冯为我们带来了另一种令人兴奋的可能性。他们发现，宇宙线——穿过太空而来、连续的轰击着大气层的基本粒子束——也有可能导致微观黑洞的产生。

宇宙线粒子最初于1912年由奥地利科学家维克多·海斯发现；90多年过去了，关于宇宙线仍有很多未解之谜。每秒钟都会有大量的宇宙线进入大气层，并产生数以十亿计的次级粒子雨，这些次级粒子会顺利地穿过你我的身体，其中的一部分有可能被分布于这个星球上的各种专用探测器观测到。但是没有人能够完全知晓组成宇宙线的粒子究竟有哪些种类（虽然我们知道它们中的绝大部分是质子），我们仅仅知道宇宙线中的一部分高能粒子来自于超新星的爆发。至于能量最高的那些宇宙线粒子究竟起源于何方，人们还没有什么好的想法。比方说，1991年10月15日，位于犹他州沙漠中的蝇眼宇宙线探测器观测到一个能量相当于300亿个质子质量的粒子划过天际。这一粒子所具有的能量如此巨大，几乎同马里亚诺·李维拉[78]投出的快球中的单个亚原子粒子所具有的能量一样大；是大型强子对撞机（LHC）上产生的粒子的能量的1亿倍。[6]这样的观测事实令人非常困惑，因为没有任何已知的天体物理过程能够产生如此高能量的粒子；实验学家们一直在用更加精确的探测器收集更多的数据以便解决这一谜题。

对于夏皮尔和冯来说，超高能宇宙线粒子究竟来自何方还不是最

值得关注的问题。这两位物理学家认识到，不论这样的粒子来自于哪里，只要微观水平上的引力远远强于人们以前所认为的程度，这些超高能宇宙线粒子就有能力在撞入高层大气的时候创造出一个小黑洞。

通过碰撞产生出来的这些小黑洞对实验学家们和大尺度上的世界完全无害。这些小黑洞一产生出来很快就会分解，随之放出大量具有某种特征的其他比较基本的粒子。事实上，微观黑洞非常短命，以至于实验学家们甚至没有办法直接探测到它们的存在；实验学家们只能通过仔细分析落到探测器上的微观黑洞粒子雨来发现蛛丝马迹。世界上最灵敏的宇宙线探测器，皮埃尔·奥格天文台——可观测的范围差不多有整个罗德岛那么大[79]——正在阿根廷西部的大草原上建造中。夏皮尔和冯估计，如果所有的额外维度都是$10^{-14}$米那么大的话，那么只要收集一年的数据，奥格探测器就有可能发现由产生于高层大气的微观黑洞导致的特征粒子碎片。如果奥格探测器没有发现这样的微观黑洞信号，那么额外维度就必须更小。找到产生于宇宙线碰撞的微观黑洞的残留信息的概率当然很小，但一旦成功，就无疑为我们打开了第一扇能看得到额外维度、黑洞、弦论以及量子引力的窗户。

除了黑洞的产生，研究人员在未来的10年间还可以利用另外一个基于加速器的方法来寻找额外维度。有的时候，兜里的硬币悄悄地就不见了，怎么回事呢？因为硬币顺着兜中的漏洞跑到衣服的夹层中了。用加速器来发现额外维度这一方法的核心思想就是复杂版的“跑到夹层中了”。

能量守恒是物理学中的一条核心原理。尽管能量可以以多种形式存在——被球棒击飞的棒球因为运动而具有动能，因为向上飞行而具有重力势能，因为撞击地面和激发各种振动而具有声能和热能；但只要你把所有种类的能量全部算清楚，你就会发现过程结束时的总能量总是等于过程开始时的总能量。[7]直到今天，人们还没有发现任何与这一完美的能量平衡定律相冲突的物理事件。

但是在考虑额外维度理论时，情况可能会有所不同。人们可能会在最新升级的费米实验室和即将运行的大型强子对撞机上的高能物理实验中发现一些破坏能量守恒的过程——碰撞结束时的能量少于碰撞开始时的能量的过程，不过，具体如何要看假想中的额外维度究竟有多大的尺寸。造成这种情况的原因有点类似于你弄丢的硬币：能量（引力子所具有的）也有可能钻到缝隙中——微小的额外空间——从而导致计算能量的时候会少掉一部分。这种“丢失能量信号”的可能性以另一种方式告诉我们，宇宙的结构所具有的复杂性远超我们的直接所见。

必须承认，我对额外维度的理论稍有些偏心。毕竟，我在这一领域奋斗的时间已经超过15年了，额外维度的某些方面在我心中占有特殊地位。不过，即使承认了我的偏心，我还是要说：我很难想象出有什么发现比找到超出了我们所有人都熟悉的三维的额外维度的证据更令人兴奋了。在我心目中，眼下还没有什么其他重要想法的实验验证能够如此彻底的震撼物理学的基础，能够使我们必须去质疑基本层面上看起来不证自明的真实性原理。

希格斯、超对称，还有弦论

近来，人们之所以要升级费米实验室的加速器和建造庞大的大型强子对撞机，并不仅仅是出于探索未知的科学好奇心以及发现额外维度的考虑，还有很多特殊的动机，其中之一就是找到希格斯粒子。我们在第9章中曾经讨论过，令人迷惑不解的希格斯粒子是希格斯场——物理上假想的场，其所形成的希格斯海能赋予其他种类的基本粒子以质量——的最小组成。当前的理论研究和实验进展都在向人们暗示，希格斯粒子的质量应该在质子质量的100～1000倍。如果希格斯粒子的质量就在这一范围的下限附近，那么费米实验室就有很大的机会在未来的几年内找到希格斯粒子。当然，如果费米实验室没能成功但是估算的质量范围还是正确的话，大型强子对撞机应该在10年之内产生大量的希格斯粒子。希格斯粒子的发现将是一个里程碑式的成

就，因为它将最终确认一种理论粒子物理学家和宇宙学家在没有任何实验证据的情况下提出了几十年的粒子的存在。

费米实验室和大型强子对撞机的另一个主要目标是发现超对称的证据。回忆一下第12章，我们曾经讨论过自旋相差1/2的超对称粒子对以及超对称的想法如何起源于20世纪70年代早期的弦论研究。如果真实世界真的具有超对称性，那么每种已知的自旋1/2的粒子都会有一种自旋0的超对称伴；每种已知的自旋1的粒子都会有一种自旋1/2的超对称伴。比如说，自旋1/2的电子会有一种自旋为0的伙伴，称为超对称电子（suPersymmetricelectron），或简称为超电子（selectron）；自旋1/2的夸克会有一种自旋为0的伙伴，称为超对称夸克（suPersymmetricquarks），或简称为超夸克（squarks）；自旋1/2的中微子会有一种自旋为0的超中微子（sneutrino）相伴[80]；对于自旋为1的胶子（gluon）、光子（Photon）、W玻色子和Z玻色子来说，也分别有自旋1/2的gluinos, Photinos, winos与zinos相伴（是的，物理学家们在命名上总有些偷懒）。

那么为什么没人探测到这些假想中的粒子呢？对此，物理学家们只能解释为这些超对称粒子的质量比对应的已知粒子的质量大。理论分析表明，超对称粒子的质量可能是质子质量的1000倍左右，如果真是这样的话，实验上没有看到任何这些粒子的信号就毫不为奇了——现有的原子对撞机没有足够的功率来制造出这些粒子。不过，这一现状将在接下来的10年间得以改变。首先，费米实验室最近升级的加速器就有可能发现超对称粒子。其次，就像前文关于希格斯的讨论一样，要是费米实验室没有发现超对称的证据，但之前的理论对超对称质量范围的估计非常准确的话，大型强子对撞机就应该能够制造出这些粒子。

超对称的实验验证将是最近这20多年间基本粒子物理领域最重大的进展。这一进展将使我们对于超出成功的粒子物理标准模型的新物理的理解更进一步，并且间接证实弦论至少没有在错误的轨道上前

进。但请注意，它并不能证明弦论本身。虽然超对称是在发展弦论的过程中建立起来的，但是物理学家们早就认识到超对称是更具普遍性意义的原理，并且可以非常容易地纳入传统的点粒子物理方案中。超对称的实验验证虽然确立了弦论体系的一个重要组成部分并且会为下一步的研究指明方向，但它绝对不能算是弦论的信号。

另一方面，如果膜世界方案正确的话，即将到来的加速器实验就有能力验证弦论。我们在第13章已经简要的介绍过，要是膜世界方案中的额外维度大到$10^{-16}$厘米的话，那么不仅引力要比以前认为的大，弦也比以前认为的要长得多。因为弦的长度越大硬度就会越小，振动弦所需要的能量也就会越小。在传统的弦论体系中，弦的振动模式所具有的能量超出了当代加速器最高能量的千万亿倍；而在膜世界方案中，弦的振动模式所具有的能量可能只有质子质量的一千倍。要真是这样的话，大型强子对撞机上的高能对撞就会像在钢琴里跳来跳去的高尔夫球一样，有足够多的能量弹奏出弦振动模式的多种音节。实验学家们将会发现大量的前所未见的新粒子——也就是大量的前所未见的弦的振动模式——这些新粒子对应着弦论中不同的谐振模式。

这些粒子的性质及其之间的关系将明白无误地告诉我们：它们都只是同一壮丽的宇宙乐章的一部分，它们虽不相同但却是彼此相关的音符，它们都只是同一种物体——弦——的不同振动模式。在可预见的将来，膜世界方案将是弦论最可能被直接验证的一种方案。

## 宇宙的起源

我们已经在前面的章节中看到，宇宙微波背景辐射在20世纪60年代中期被发现后，就一直在宇宙学的研究中扮演着重要角色。原因很明显：当宇宙处于幼年的时候，空间中满是带电粒子——电子、质子，等等——这些带电粒子会由于电磁力的缘故而连续的辐射光子。到了大爆炸之后的30万年左右，宇宙逐渐冷却，电子和质子组成了电中性的原子；从这个时候开始，辐射就开始几乎不受干扰的在整个空

间中穿行，从而为我们留下早期宇宙的快照。每立方米的空间中差不多有4亿个原初宇宙微波光子，它们就是早期宇宙留下的遗迹。

最早测得的宇宙微波背景辐射在温度上呈现出明显的均匀性。但正如我们在第11章中讨论过的那样，晚近的一些探测——最早的宇宙背景探测器（COBE）以及稍后的一系列更先进的探测器所做的一些探测——发现了一些温度细微改变的证据，如图14.4（a）所示。图中，不同的灰度标示着不同的数据，较亮的部分和较暗的部分之间的温度差一般在万分之几度左右；那些斑点表示的是天际中微小但却不可忽略的温度变化。

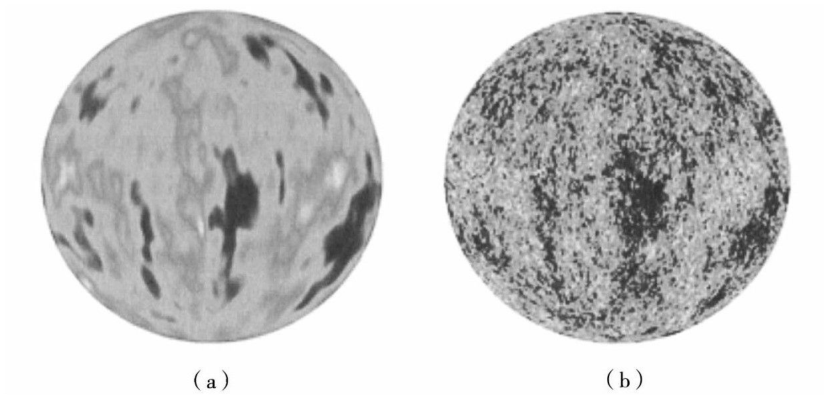



图14.4　（a）：COBE卫星所收集的宇宙微波背景辐射。自从大爆炸后的30万年起，这种辐射就无阻的穿行于宇宙中，所以这张图片反映的是距今差不多140亿年以前的宇宙微小温度变化。（b）：由WMAP收集到的更加精确的数据。

COBE实验不但有了重大的发现，还从根本上改变了宇宙学研究的特点。COBE之前的宇宙学数据一般非常的粗糙。那个时候，一个宇宙学理论只要能够大体上符合天文学观测，就会使人们相信它。理论学家们可以在几乎不怎么需要理会观测限制的情况下抛出一个又一个的理论。因为本来观测限制就少得可怜，仅有的几个又非常的不精确。但是COBE开启了一个新纪元，宇宙学理论要受到一系列标准的严格限制。现在，任何一个新提出的理论在被人们接受之前，先得成

功的算出还在不断增加的大量精确的实验结果。2001年，由NASA和普林斯顿大学合资兴建的威尔金森微波各向异性探测器（WMAP）卫星开始以近乎COBE的40倍的分辨率和灵敏度测量微波背景辐射。将WMAP的最初结果[图14.4（b）]与COBE的结果[图14.4（a）]相对照，你就会立即看出WMAP能够画出的图究竟有多么精细。正在由欧洲空间局建造中的另一颗卫星——普朗克——计划于2007年发射，如果一切按计划进行，普朗克卫星的分辨率将达到WMAP的10倍左右。

大量精确实验数据的出现结束了宇宙学研究中良莠不齐的局面，暴涨理论成了主要的理论候选者。但是，我们在第10章曾经提到过，暴涨理论并不是只有一个版本。理论学家们已经提出了很多个不同版本的暴涨理论（旧暴涨理论，新暴涨理论，暖暴涨理论，混合暴涨理论，超暴涨理论，援暴涨理论，永暴涨理论，扩充暴涨理论，混沌暴涨理论，双暴涨理论，弱标度暴涨理论，超自然暴涨理论，要知道这些还不是全部），每一种版本都具有标志性的短时间急速膨胀爆发阶段，但是在细节上又各不相同（场的数目或势能的形状有所不同，以及究竟是哪一种场位于势能最低的位置等区别）。这些理论中的差别导致了在预言微波背景辐射性质时候的区别（具有不同能量的不同场有些微不同的量子涨落）。通过与WMAP和普朗克卫星所得到的实验数据相比较，我们可以排除很多种理论上的可能性，使我们对宇宙的理解更进一步。

事实上，我们可以利用实验数据来为宇宙学研究领域进一步瘦身。尽管被暴涨膨胀放大的量子波动可以为观测到的温度变化提供一个合理的解释，但是暴涨理论还是有一个竞争者。我们在第13章讲过的由斯坦哈特和塔洛克提出的循环宇宙模型就是另一种可能的理论候选者。当循环宇宙模型中的两个3-膜彼此相对靠近的时候，量子涨落会使不同的部分以些微不同的速率彼此接近。当这两片3-膜在差不多3万亿年后最终撞在一起的时候，膜上的不同位置会在不同的时刻彼此碰撞，就像两张粗糙的砂纸拍在一起那样。两片膜没能够完美的均

匀接触导致了每一片膜不能完美的均匀演化。而我们已经假定这两片膜中的一片就是我们的三维空间，所以膜的非完美均匀演化就是我们能够探测到的不均匀性。斯坦哈特、塔洛克及其合作者提出，这种不均匀性导致的温度变化可以与暴涨理论所预言的温度变化具有相同的形式；因此，只用我们现在所拥有的数据的话，我们没法区分循环宇宙模型与暴涨理论的宇宙学预言，两者都能解释当前的实验观测。

不过，在未来的10多年里，越来越多的精细数据有可能将两种方法区分开。在暴涨理论的框架下，被指数膨胀放大的并不只有暴涨子场的量子涨落，还有空间结构中的微小量子波纹。因为空间中的波纹不是别的，正是引力波（参见我们关于LIGO的有关讨论），所以暴涨理论预言早在宇宙的最初时刻就有引力波产生。[8]我们一般将这种引力波称为原初引力波，以区分于晚近时期由于猛烈的天体物理现象而产生的引力波。而在循环宇宙模型中则正好相反。在这一模型中，对均匀性的偏离是慢慢建立起来的，整个过程所用掉的时间长得不可想象，这是因为两片膜要花差不多3万亿年的时间才能碰撞一次。膜的几何结构和空间的几何结构并没有迅速的改变，这意味着根本就不会出现空间波纹。因而，循环宇宙模型根本就没有预言原初引力波的存在。所以，一旦原初引力波被实验验证，那就意味着暴涨理论又取得了一次重大胜利，而循环宇宙模型则被实验排除。

LIGO的灵敏度很有可能无法探测到暴涨理论所预言的引力波，但是建造中的普朗克卫星和另一个卫星实验——宇宙微波背景偏振实验（CMBPol）——则有可能实验验证这一预言。这两个实验，特别是CMBPol，所关注的并不仅仅是微波背景辐射的温度变化，还会测量偏振，所探测到的微波光子的平均自旋方向。详细解释起来会涉及一连串的相关知识，所以我们在这里只是简单地说一下：来自于大爆炸的引力波可能会在微波背景辐射的偏振中留下某种印记，而这种印记可能大到足以被实验发现的程度。

所以，10年之内，我们就可能会搞清楚究竟是真的有大爆炸这么

一回事呢，还是我们所熟悉的宇宙实际上是一张3-膜。在这个宇宙学的黄金年代，即使一些最疯狂的想法也得到了实验检验的机会。

暗物质、暗能量以及宇宙的未来

我们曾在第10章中了解到：大量的理论和观测证据表明，宇宙的组成中只有5%是我们熟悉的物质——质子和中子（电子在普通物质中所占的份额少于0.5%）——25%是所谓的暗物质，而另外的70%是暗能量。但是，物理学家们仍然没能搞清楚这些暗物质和暗能量究竟是些什么。很自然的，人们首先会猜想暗物质也是由质子和中子组成的，只不过以某种特殊的方式组合在一起，没有形成发光的星体。然而，理论上的原因使得这样的猜想不可能正确。

通过精细的实验观测，天文学家很清楚整个宇宙中到处都是轻的核元素——氢，氦，氘，锂——的平均相对丰度。物理学家们相信这些轻核通过某一过程形成于宇宙的最初几分钟，理论计算这种形成过程得到的结果与实验观测符合得非常好。理论和实验的这一精确符合是现代理论宇宙学的重大成就之一。但是，这种理论计算首先假定暗物质不是由质子和中子组成的；如果暗物质是由质子和中子组成的话，那么在宇宙的尺度上，质子和中子就会成为宇宙的主要构成物质，这样一来实验观测就将理论排除掉了。

那么，如果组成暗物质的不是质子和中子，又是什么呢？直到今天，虽然人们提出了大量的可能性，可是还没有人能够真正解决这个问题。从轴子到zino的很多名字都被人们拿出来当暗物质的候选者，毫无疑问，任何一位回答了这一问题的科学家必将被请到斯德哥尔摩一游[81]。人们还未曾探测到哪怕是一个暗物质粒子，这一事实对暗物质的候选者们提出了很强的限制。这是因为暗物质并不只存在于外太空，它们遍布于整个宇宙，也会存在于你我的身边。关于暗物质的很多理论都会告诉我们，每秒钟都会有数以10亿计的暗物质粒子穿过你我的身体，因而可能的暗物质候选者必须得是那些穿过物质但却不留

下痕迹的粒子。

中微子可算是一种可能性。计算表明，大爆炸产生出来的中微子的残留丰度大约是每立方米5500万，只要3种中微子中能有一种重达质子质量的一亿分之一（10-8），它们就能够被当做暗物质的候选者。尽管近来的实验已经获得了中微子具有质量的有力证据，但是测得的中微子质量实在太小，比所需要的程度小了差不多100倍，所以中微子很难是暗物质。

另一个比较有希望的提议与超对称粒子——特别是Photino， zino以及higgsino（分别是光子、Z玻色子和希格斯粒子的超对称伴）——有关。上述的这些粒子是超对称粒子家族中最冷漠的一些家伙，它们常常可以在几乎不受影响的情况下毫无声息地穿过地球，使得我们很难追寻到它们的踪影。[9]通过计算在大爆炸过程中到底产生了多少这样的粒子以及存活到今天的还有多少，物理学家们估算出这些粒子的质量应该在质子质量的100～1000倍之间，只有这样它们才能充当暗物质。这是一个非常诱人的结果，因为人们在完全不考虑暗物质和宇宙学的情况下，单从超对称粒子模型和超弦的各种研究中得出的相关粒子质量范围也是这么大。两类研究的结果交汇到了一起，这样的事情只有在暗物质就是由超对称粒子构成的情况下才能说得通。因而，我们也可以把在世界上现有的和即将启用的加速器上寻找超对称粒子看做是寻找可能性很高的暗物质候选者。

直接探测穿越地球而过的暗物质粒子的实验也进行一段时间了。毫无疑问，这样的实验极具挑战性。每秒钟，在1/4平方米的面积上大约会穿过100万个暗物质粒子；但即使这样，每天能在专为暗物质而设计的探测器上留下痕迹的暗物质粒子一般不会超过一个。到目前为止，人们还没能成功地探测到暗物质粒子。[10]既然暗物质还没有使任何人获得诺贝尔奖，实验学家们当然会更加努力迎难而上。在未来的几年间，暗物质的身份很有可能最终得以确认。

暗物质存在的最终确认及其身份的直接认证将是科学上的重大进步。人类将有史以来第一次搞清如此基本却又难以捉摸的东西：宇宙的主要物质组成。

正如我们在第10章中看到的那样，近来的实验数据强烈地向我们暗示，即使暗物质的身份得以确认，有关宇宙内容的版图中仍有一大块尚需实验检审：对超新星的观测提供了一些证据——宇宙中70%的能量可用一种具有外推作用的宇宙常数来说明。作为过去10年间最令人激动并且最出乎意料的发现，宇宙常数——充塞于空间的能量——的证据还需要更为严格的检验。为此，人们已经想出了很多的办法，一些还在计划之中，而另一些已经启动。

微波背景辐射实验也要扮演非常重要的角色。图14.4中的斑点——每一个斑点代表的是温度相同的一块区域——反映的是空间结构的整体形状。如果空间的形状像球一样，比如说像图8.6（a）所示的那样，向外的膨胀就会使斑点变得比图14.4（b）中的大一点；如果空间的形状像图8.6（c）所示的那种马鞍面，向内的收缩就会使斑点变小一点；如果空间的形状像图8.6（b）所示的一样是平面，斑点就有可能变大也有可能变小。由COBE首先进行，WMAP进一步改善的精确测量强有力地支持了空间是平坦的这一主张。这样的测量结果不仅与暴涨模型的理论预言相吻合，也与超新星的观测结果完美吻合。我们已经知道，宇宙的空间平坦意味着总质量或总能量密度等于临界密度。这样一来，普通物质和暗物质一共占宇宙总密度的30%，暗能量再贡献了余下的70%，一切就都和谐一致了。对超新星结果的进一步直接确认是超新星加速探测器（SNAP）的目标之一。由劳伦斯·伯克利实验室的科学家们企划的SNAP是一台随卫星轨道运动的望远镜，它将观测的超新星数目是目前已研究过的数目的近20倍。SNAP不仅能确认早前的观测结果——即宇宙的70%为暗能量，还将更加精确的测定暗能量的性质。

你瞧，虽然我把暗能量描述成了爱因斯坦宇宙常数的另一个版本

——恒定不变的推动着空间膨胀的能量——但还是有另一种密切相关却有所不同的可能性。还记得我们有关暴涨宇宙学的讨论吗（那只四处蹦的青蛙）？某个场在其场值高于最低能量时可以像宇宙常数一样，驱使空间加速膨胀，不过这样的过程仅能持续一小段时间。这个场迟早会回归到其势能碗的最低位置，向外的推力也随之消失。在暴涨宇宙学中，这个过程发生于短短的一瞬间。但要是引入一种新的场并小心的选取其势能形状，物理学家们就有办法使加速膨胀变得不那么猛烈但却更加持久，这样的话，该场就可以在跌回到最低能量位置之前，以相对较慢但却持久的外推力驱动空间加速膨胀很长时间——长达几十亿年。这样的想法引出了另一种可能，即，我们有可能正在经历极度柔和版的暴涨膨胀——且有理由相信这一膨胀过程开始于宇宙的最初时刻。

真实的宇宙常数与后一种可能性——即所谓的精质（quintessence）——之间的区别对于今天来讲并不重要，但从长远来说却对宇宙影响深远。宇宙常数是一个常数，它使得宇宙可以永不停息的加速膨胀。在宇宙常数的作用下，宇宙的膨胀会变得越来越快，宇宙的疆域也会越来越辽阔，同时，宇宙也变得更加稀薄，荒凉。但是由精质导致的膨胀会在某个时刻之后慢慢终结，与永远加速膨胀的宇宙相比，这样的宇宙拥有一个不那么荒凉的未来。通过测量空间加速度在长时间间隔的变化（通过观测不同远近——也就是不同时间——的超新星来进行这种测量），SNAP将有可能得以辨别这两种可能性。一旦SNAP为我们解开暗能量是否真的就是宇宙常数这一谜题，我们就有机会洞察宇宙未来的命运。

## 空间、时间以及猜想

探索空间和时间性质的旅程漫长遥远，其间满是各种惊奇，而且毫无疑问，人类在这一旅程中仍处于起步阶段。在过去的几个世纪里，人类经历了一个又一个的重大突破，这些突破以激进的方式一次又一次的改变了我们关于空间和时间的概念。我们在这本书中所讨论

的理论和实验进展代表着我们这一代的物理学家们对这些概念的梳理，并且很有可能就是我们科学遗产的主要部分。在第16章中，我们将要讨论一些最新的带有猜想性质的进展。通过这样的讨论，我们或许可以看一看人类探索旅程的下一步可能通向何方。但是首先，我们将在第15章中想想另外一些不同的方向。

科学发展没有确定的模式，历史一再地告诉我们，思想上的突破通常是通往技术手段的第一步。人类在19世纪理解了电磁力，正是这一理解使我们最终拥有了电报、无线电和电视。有了电磁的相关知识，再加上稍后人类对量子力学的理解，我们又拥有了计算机、激光，以及种种数也数不过来的电子器件。对核力的理解既使人类得到了有史以来最强大的危险武器，也使人类有希望在未来的某一天只靠大量的盐水就能满足整个世界的能源需求。我们对空间和时间的深入理解会不会也只是类似的技术发展模式的第一步呢？我们有没有可能在未来的某一天了解时间和空间的奥妙，利用我们的相关知识实现一些现今只能出现在科幻故事中的构思呢？

没有人知道，但是我们可以一起看看我们已经有了些什么以及哪些神奇的构思有可能在未来的某一天得以实现。

# 第15章　超距传输器与时间机器

## 在时空中旅行

退回到20世纪60年代，当时的我或许真的是缺乏想象力。但在企业号[82]的甲板上看到电脑确实令我感到难以置信。作为一个20世纪60年代的小学生，我可以接受空间跃迁，我也可以接受宇宙中到处都是说着英语的外星人；但我真的难以想象竟然有一台这样的机器：它可以根据要求立即播出历史人物的画面，详细解说任何已有设备的技术细节，又或者调出任何一本已出版著作。这样的一台机器超越了我的想象极限，令我很难相信。20世纪60年代末的时候，一个小孩当然会

认为永远都不会有办法收集、存储如此巨量的信息。但仅仅半个世纪后，我就可以坐在厨房里用笔记本电脑无线上网，可以使用语音识别系统，还可以看《星际迷航》，手都不抬一下就可以在巨型的知识库中寻找资料——重不重要的都可以找到。诚然，《星际迷航》中的23世纪电脑有着令人羡慕的速度和效率，但是今天的我们也可以预见，一旦真的到了23世纪，我们的电脑技术将大大超越影片中勾画的水平。

上面讲的只是科幻小说预言未来的众多例子中的一个。但在《星际迷航》这部电视剧中，最值得称道的仪器还得算是超距传输器——走进一间舱室，按一下按钮，然后你就被传送到遥远的地方或完全不同的时代。有没有可能在未来的某一天，人类真的可以超越空间与时间的局限，自由穿梭于时空，探索时空的最远疆界呢？科幻小说与科学之间的鸿沟有没有可能被填平呢？考虑到我已经告诉过你们，我小的时候完全没有办法相信真的会有信息革命到来的这一天；你们完全可以质疑我在预言未来技术突破方面的能力。所以，我们在这一章中不会妄加猜测未来会有什么，而要谈谈在朝着掌控空间和时间、实现超距传输器和时间机器前进的方向上，我们在理论和实践上已经取得了哪些进展。

量子世界的瞬间移动

在传统的科幻故事中，超距传输器（或者按照《星际迷航》中的名称，传送器）先要扫描某个物体以确定其全部组成信息，然后将这些信息发送到远方的某个位置，在那里，另一台机器将按照这些信息重构该物体。不管是先将物体本身“分解”，然后将其原子分子与蓝图一起传送到远处来构建该物体的副本，还是直接用远端的分子和原子来构建物体的副本，都只是不同版本的小说式虚构。我们将会看到，过去10年间发展起来的超距传输方法在本质上与后一种情形倒有些接近，但由此引出两个问题。第一个是标准但却棘手的哲学难题：如果真的可能的话，究竟从什么时候开始，我们才可以将副本识为、称

为、认为是原始的物体，并像对待原始物体一样对待副本？第二个问题是是否有可能——即使只是理论上的可能——完美地扫描一个物体，准确地探明其组成成分以便我们可以完美的绘制出该物体的蓝图从而重建该物体。

在由经典物理定律掌控一切的宇宙中，我们对第二个问题可以做出肯定的回答。理论上，组成一个物体的每个粒子的所有性质——每个粒子的类型、位置、速度，等等——都可以完全确定下来，并作为重构物体的蓝图传送到远方。当然，完全确定组成一个物体的全部基本粒子的所有信息会难得超乎想象；但是，在经典宇宙中，唯一的障碍来自于复杂程度，而不是物理。

在一个由量子物理掌控的宇宙中——比如说我们的宇宙就是这样，情况则不是这么简单。我们已经知道，所谓的测量将使一个物体种种可能的性质中的一个脱离量子迷雾，使之获得确定的值。比如说，当我们观测一个粒子的时候，我们所观测到的当然是某一确定的性质，但这一性质并不能反映我们观测之前该粒子所具有的杂烩式量子性质。[1]因而，一旦我们想要复制一个物体，我们就将面对量子的第二十二条军规[83]。要想复制我们就必须知道要复制些什么，要想知道复制些什么我们就必须观测，而观测又会造成改变，所以我们要是按照我们所看到的进行复制的话，那复制的产物就不是观测之前的那个物体了。这就表明在量子世界中，超距传输是不可能实现的，并且这种不可能并不是由技术上的复杂性造成的，而是由量子物理的先天局限性造成的。但是，我们在下一节中将会看到，20世纪90年代早期，一个国际物理学家团队找到了一种巧妙的方法绕开了这一结论。

至于第一个问题，即原始物体与副本之间的关系，量子物理给了一个明确又鼓舞士气的答案。根据量子力学的原理，宇宙中的所有电子都彼此类同，因为它们都具有完全一样的质量，完全一样的电荷，完全一样的弱核力和强核力性质，以及完全一样的自旋。而且，已经经受住了实验检验的量子力学告诉我们：上面所列举的这些电子性质

就是电子所能具有的全部的性质。按这些性质来看，全体电子彼此类同；而且也不存在其他可以用来区分电子的性质。同样，所有的上夸克彼此全同，所有的下夸克全同，所有的光子全同，总之，任何一种基本粒子都会彼此全同。几十年前量子方面的先驱者就认识到，粒子可以被看做是一个场最小可能的波包（比如说光子就是电磁场最小的波包），而且，根据量子力学，一个场的这种最小组成总是全同的（或者，我们可以在弦论的理论框架下这样理解，同一种类的粒子之所以有全同的性质是因为它们都是同一种弦的全同振动模式）。

同一种类的两个粒子唯一有可能有所区别的地方是它们处于不同位置的概率，它们的自旋指向特定方向的概率，以及它们具有特定的速度和能量的概率。又或者按照物理学家们习惯的说法，两个全同粒子可以处于不同的量子态。但要是同一种类的两个粒子处于同一种量子态的话——有一种可能性不能算在内，即，一个粒子有极大的概率在这，而另一个粒子有极大的概率在那——量子力学原理就会保证它们不可区分，并且这种不可区分并不仅是实践意义上的，更是理论意义上的。这样的粒子可算是完美的双胞胎。一旦两个粒子交换彼此的位置（或者更准确地说，交换两个粒子处于给定位置的概率），我们将没有任何办法发现这种交换。

因此，我们可以这样想，开始的时候我们把一个粒子放于此处，
84然后不管通过什么办法把另一个放在远处的同一种类粒子置于完全相同的量子态（使之具有相同的自旋指向概率、能量概率等），这样制备的粒子就将与原始粒子不可区分，这样的过程就可以称为量子超距传输。当然，要是原始粒子在整个过程中毫发无损的话，你可能更愿意将这个过程称为量子克隆或量子传真。但是我们将会看到，这些想法的科学实现将无法保护原始粒子——在超距传输过程中它将会不可避免的被改变——所以我们不会为到底取什么名称而处于两难。

很多哲学家以不同的方式思考过的一个更为紧要的问题是，在一个粒子身上能实现的事情是不是也能在真正的宏观物体身上实现呢？

如果你可以将你的DeLorean[85]的每一个组成粒子都从一个地方传输到另一个地方，并且在这个过程中确保每个粒子的量子态以及彼此之间的相互关系100%的被复制，那你是不是就成功的传送了一台轿车呢？尽管没有实践经验可供参考，但是理论上得来的证据倒是强烈的支持已经成功传送这样的结论。决定一个物体看起来是什么样子，摸起来是什么感觉，听起来是什么声音，闻起来甚至尝一下是什么味道的就是物体中原子和分子排列，所以传送过去的轿车应该就是原始的DeLorean——碰花的地方还在那里，左边的车门还是嘎吱嘎吱的响，你养的狗留下的尿骚味什么的也全都有——它也能像原来的那辆一样随时急转弯，油门踩起来的感觉也不会有所不同。传送过去的车究竟是原来的那辆还是精确的副本这一问题无关紧要。如果你要求联合量子海陆货运公司[86]将你的轿车用轮船从纽约运往伦敦，但他们却悄悄地用了超距传输的办法传送过去，那么只凭辨认的话你永远也不会知道他们没按你的要求做——甚至连理论上的可能性都没有。

但搬运公司传送的是你的猫，或者为了满足你那独特的品位，你要求搬运公司对你本人来一次越洋传送，那又会有什么问题呢？走出接受室的猫或者人还是走进超距传输器的那只猫或那个人吗？我个人认为，是的，猫还是那只猫，人还是那个人。再次声明，我们没有任何相关数据，我或者任何其他人能做的都只是猜测。但是按我的思考方式，任何一个活着的人，只要他体内的全部原子和分子与组成我身体的原子和分子处于完全一样的量子态的话，那我就要说“他”就是我。即使“原始”的我在“拷贝”生成后仍然存在，我（我们）也会毫不犹豫的宣称每一个都是我。我们应该有同样的想法，都会发自肺腑的觉得彼此并不高于对方。思想、记忆、情感和看法这些东西建立在组成人体的分子与原子性质的基础上；要是这些基本成分具有相同的量子态的话，那么由这些基本成分构成的人也会有完全一样的意识。时光流逝，我们各自的经历将使我们彼此不同；但是我相信，从此以后将有两个我，而不是一个“真一点”的原始我加上一个“假一点”的拷贝我。

事实上，我倒愿意不那么严格的讨论一下。我们的物理组成无时无刻不在变化，只不过有的时候变得多些，有的时候变得少些，但是我们还是我们自己。哈根达斯冰淇淋会使我们血液中的脂肪和糖的含量增多；MRI（磁共振）会使大脑中一些原子核的自旋方向改变；心脏移植和抽脂术自不必说；每一百万分之一秒，普通人身体中会有一万亿个原子焕然一新，我们处于连续不断的变化之中，但是我们每个人的身份并没有发生变化。所以，即使超距传输后的那个“我”与本来的我在物理态上并没有完全吻合，那个“我”和本来的我仍可能是一模一样的人。在我的书中，那个“我”完全可以成为真的我。

当然，如果你相信除了物理成分，生命还意味着很多其他的东西的存在，特别是心灵的话，那么你的超距传输的成功标准可能会比我严格一些。人们关于这一棘手的问题——我们每个人的身份究竟在多大程度上取决于我们物理上的身体——已经以各种不同的形式争论过多年，但还是没有找到令所有人满意的答案。我认为一个人的身份只取决于其物理上的身体，而另外一些人则并不认同。总之，没有人可以宣称已经找到了终极答案。

我们姑且先不讨论你在传输人类这一假想问题上究竟持何种观点。借助于量子力学的神奇力量，科学家们已经成功的证明单个粒子可以——实际上已经实现过——超距传输。

我们一起来看看。

量子纠缠与量子传输

1997年，其时还在因斯布鲁克大学的安东·泽林格领导的一组物理学家和罗马大学[2]的A.弗朗塞斯科·德·玛蒂尼领导的另一组物理学家分别成功地实现了光子的超距传输。在这两个实验中，处于某一特别的量子态的初始光子被成功的传送到了另一个位置，虽然这两次实验只是横跨实验室的短距离传输，但是人们有理由相信同样的过程可以

在任意距离上实现。这两个实验组所使用的技术基于另一组物理学家——IBM沃森研究中心的查尔斯·本耐特；蒙特利尔大学的吉尔斯·布拉萨德，克劳德·克莱玻和理查德·约茨扎；以色列物理学家艾舍尔·帕里茨；以及威廉斯学院的威廉·伍特斯——1993年关于量子纠缠的理论探讨（参见第4章）。

回想一下，两个处于纠缠状态的粒子——比如说两个光子——有一种奇特又密切的关系。若每一个光子都有确定概率的自旋指向（向下或向上），并且每一个光子在被测量的时候都会在各种可能性中随机“选择”，那么一旦这两个光子中的某一个做出“选择”，另一个就会跟着立即做出“选择”，即便空间间隔很远也是如此。我们曾在第4章中说明了人们无法利用纠缠粒子来实现两个不同位置之间的信息超光速传输。要是我们在相隔很远的位置上分别放置纠缠光子，然后分别连续测量这些纠缠光子，那么我们就会发现每一台探测器所收集到的数据只是一些随机序列（与粒子的概率波相一致的粒子自旋指向）。只有当我们对比不同的探测器上收集到的结果时，我们才会清楚地看到这些结果惊人的一致。但要实现这种对比的话，我们必须首先通过某种常规的、低于光速的通信方式交换彼此所得到的结果。既然进行结果对比之前人们无法获知任何可以表明不同位置的光子处于量子纠缠的证据，人们实际上就没法通过量子纠缠实现超光速通信。

不过，即使纠缠现象无法用于实现超光速通信，人们还是会形成这样的感觉——粒子之间的长距离关联非比寻常，很有可能会在某些非常规的事情上起作用。1993年，本耐特及其合作者就发现了这样一种可能性。他们发现量子纠缠很有可能可用于量子传输。你可能实现不了超光速通信；但如果你想要的只是低于光速的粒子传输的话，量子纠缠可能会帮上你的忙。

这一结论背后的推理巧妙曲折——尽管数学上直截了当，我们现在就来感受一下。

假设我现在要把一个光子——称为光子A——从我纽约的家中传送到我在伦敦的朋友尼古拉斯那里。简单起见，我们来看一下我如何才能够准确的传送光子A的自旋所处的量子态，也就是说，我怎样才能使尼古拉斯获得一个自旋指向概率与光子A的自旋指向概率完全相同的光子。

我不能先测量一下光子A的自旋，然后打电话告诉尼古拉斯，让他在伦敦制备一个与我观测到的光子一样的光子；因为我观测到的结果将会被我的观测本身影响，所以观测后的结果不能真实地反映观测之前光子A所处的状态。那么我能怎么做呢？本耐特及其合作者提出，第一步是把一对处于纠缠态的光子——光子B和光子C——分给我和尼古拉斯，人手一个。我们怎样获得这两个光子并不重要。我们就先假定大洋彼岸的我和尼古拉斯分别得到了这对光子中的一个，比如说我得到了光子B而尼古拉斯得到了光子C，那么要是我沿着某一给定轴测量光子B的自旋，尼古拉斯也对光子C做同样的测量的话，我们就将得到完全一样的结果。

然后第二步，根据本耐特及其合作者的步骤，并不是直接测量光子A——我想要传输的那个光子——那样明显没有什么干涉作用。我应该做的是测量光子A及纠缠光子B的联合性质。例如，根据量子力学原理，我可以在不分别测量每个光子自旋的情况下测量光子A和光子B是否关于某一垂直轴具有相同的自旋。与此类似，量子力学也允许我在不分别测量每个光子自旋的情况下测量光子A和光子B是否关于某一水平轴具有相同的自旋。虽然我无法利用这种测量得到光子A的自旋，但我却能测得光子A的自旋与光子B的自旋之间的关系。而这种关系是非常重要的信息。

远在伦敦的那个光子C与光子B处于纠缠状态，所以，一旦我知道了光子A和光子B之间的关系，我就能推求出光子A和光子C之间的关系。如果我在这个时候打电话给尼古拉斯，告诉他光子A的自旋相对于光子C的自旋是怎样的，他就能知道怎样操控光子C才可以使其量子

态与光子A的量子态正确的匹配。在尼古拉斯实施了必要的操作之后，他所拥有的那个光子就与光子A全无二致了。然后我们就可以宣布已经成功的传输了光子A。比如我们来看一下最简单的例子：一旦我在测量后发现光子B的自旋同光子A的自旋完全一样，那我就能推断出光子C的自旋也和光子A的自旋完全一样，于是，什么都不用做了，我们已经完成了对光子A的传输过程。光子C与我们要传输的光子A处于完全相同的量子态。

大体上就是这样。上面介绍的是大体上的想法，要想解释清楚具体操作步骤上的量子传输，我必须再讲一讲一个到目前为止还没有谈及的关键要素。当我对光子A与光子B进行联合测量的时候，我测得的是光子A的自旋相对于光子B的自旋是怎样的。但是，正如在所有的量子测量中都不可避免的那样，测量本身会对光子有影响。因此，我所测得的并不是测量之前光子A的自旋与光子B的自旋之间的关系，而是在光子A和光子B都被测量行为干扰了之后两者之间的关系。所以，乍看之下，我们似乎再次遭遇了我在刚开头讨论如何直接复制光子A时遇到的量子麻烦：测量过程导致的不可避免的干扰。好在我们还有光子C。既然光子B与光子C处于量子纠缠状态，那么我使纽约的光子B受到的干扰也会在身处伦敦的光子C的身上体现出来。这就是我们在第4章中讨论过的量子纠缠的奇妙性质。事实上，本耐特及其合作者在数学上证明了测量所导致的干扰可以通过光子B和光子C之间的纠缠显现在远方的光子C身上。

这样的结果相当有趣。通过测量，我们可以知道光子A与光子B的关系，但测量也会带来一个棘手的问题——光子A和光子B都会由于测量的介入而被干扰。但是因为有纠缠的存在，光子C被引入我们的问题中——即使相差千里，量子纠缠也会起作用——并且可以帮助我们分离出干扰的效应，从而使我们获得在测量过程中丢失的原始信息。如果我现在再给尼古拉斯打电话告知测量的结果，他就知道了测量产生干扰后光子A的自旋和光子B的自旋之间的关系，然后，再通过光子

C，他就会知道干扰本身的影响。这个时候尼古拉斯就可以利用光子C除去测量所导致的干扰效应，从而跳过量子障碍，实现光子A的复制。本耐特及其合作者详细地证明了，至多通过对光子C进行一些简单的操控（如何操控要看我在电话里告诉了尼古拉斯哪些有关光子A和光子B的关系的信息），尼古拉斯就可以用光子C及其自旋指向精确的复制出在我进行测量之前的光子A所处的量子态。而且，不但光子A的自旋可以被复制，光子A的量子态的其他性质（比如说光子A处于某一能量的概率）也可以通过类似的办法得以复制。因此，利用这个方法，我们就可以实现光子A从纽约到伦敦的超距离传输。[3]

如你所见，量子超距传输包括两个步骤，其中的每一个都会传递重要但又互补的信息。首先，我们将要进行传输的光子和某一纠缠光子对中的一个光子合在一起进行联合测量。在测量过程中产生的干扰会由于古怪的量子非定域性而体现在另一个位于远方的纠缠光子身上。这就是第一步，超距传输过程中清楚的量子部分。在第二步中，测量结果本身会通过较为传统的方式（电话、传真、e-mail，等等）传送到远方的接收站，这个步骤可称为超距传输过程中的经典部分。将这两个步骤结合到一起，通过对位于远方的另一个纠缠光子进行某种直接操作（比如说绕某个特定轴进行旋转），我们就可以复制出与想要传送的光子具有相同的量子态的光子，从而实现超距传输。

有两个关键的量子传输性质需要注意一下。既然光子A的原始量子态已经被测量过程破坏，那么位于伦敦的光子C就成了唯一一个具有原始光子态的光子。原始光子并没有两个拷贝，所以我们不应该将这个过程称为量子传真，而应该称为量子传输。[4]而且，即使我们把光子A从纽约传送到了伦敦——伦敦的那个光子已经与原始的光子A完全一致——我们还是不知道光子A的量子态。现在，伦敦的这个光子已经与在我们开始种种操作之前的那个纽约的光子A有完全一样的自旋指向概率了，但是我们并不知道这个概率究竟是多少。事实上，这正是量子传输中的绝妙之处。测量导致的干扰使我们无法获知原始光

子A所处的量子态，但是按我们讲的办法，我们并不需要知道一个粒子的量子态就能传输这个粒子。我们需要知道的只是其量子态的某个方面，也就是在与光子B的联合测量中所获知的那些信息。接下来的事情就可以统统交给处于量子纠缠状态的远方光子C了。

按照这个办法实现的量子传输取得了丰硕的成果。20世纪90年代早期，制备纠缠光子对已是标准化的程序了，但人们还没有实现对两个光子的联合测量（就是前面讲的对光子A和光子B的联合测量，术语上讲就是贝尔态测量）。泽林格和德·玛蒂尼所领导的这两个实验组的成功之处就在于独创性地发明了联合测量的实验技术并在实验室中实现了这种技术。[5]1997年，这两个组分别实现了目标，成为世界上最早实现单个粒子超距传输的实验组。

现实中的超距传输

你、我、DeLorean以及所有的一切都是由很多个粒子构成，所以很自然的，下一步需要考虑的就应该是如何将量子超距传输应用到这样大的粒子集合上，从而使我们实现将宏观物体从此地传送到彼处。但是，从传送一个粒子过渡到传送整个宏观物体可不是那么简单的事，这中间要完成的很多任务都不在研究者们现阶段所具有的能力范围之内，本领域的很多权威人士甚至认为或许等相当长的一段时间后人们才有可能实现这样的目标。下面纯粹为了凑趣，我们可以一起来看看泽林格的梦想在未来将如何实现。

假设我要把我的DeLorean从纽约传送到伦敦。这次我和尼古拉斯需要的就不仅仅是一对纠缠光子了（那是传送一个光子时我们所需要的全部），现在，我们每人要有一间库房，里面装满了质子、中子、电子以及其他的一些粒子，这些粒子要多到足以建构起一辆DeLorean；而且，在我的库房中的所有粒子要和尼古拉斯的库房中的所有粒子处于量子纠缠状态（如图15.1所示）。我还需要一台设备用以对组成我的DeLorean的所有粒子和库房中飞来飞去的粒子来一次联

合测量（就好像对光子A和光子B进行联合测量一样）。因为两个库房中的粒子存在纠缠，所以我在纽约所做的测量也会影响尼古拉斯在伦敦的库房中的粒子（就如同光子C可以反映出对光子A和光子B所做的联合测量）。然后我就打电话告诉尼古拉斯我的测量结果（这次电话注定要花掉大笔的电话费，因为我要告诉尼古拉斯的是$10^{30}$个结果），这样的话，他就可以根据这些数据对他的库房中的粒子进行种种操作（就像我前面的电话告诉尼古拉斯如何对光子C进行操作一样）。一旦尼古拉斯完成了这些任务，在他的库房中的粒子就会和未测量前的DeLorean中的粒子处于完全一样的量子态，这时，就像我们之前讨论过的那样，尼古拉斯已经有了DeLorean。[87]这台DeLorean已经被成功地从纽约传送到了伦敦。

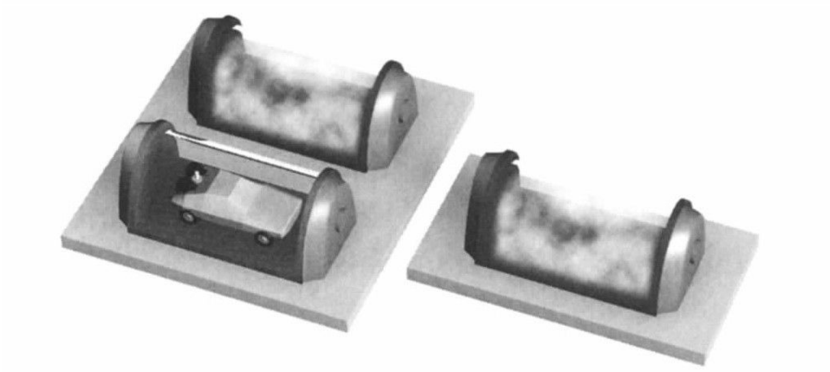


图15.1　假想中的物体超距传输。首先需要在传输地和目的地准备两个完全一样的粒子库，这两个粒子库中的粒子处于量子纠缠态；然后对传输地的粒子库和要传输的物体做一次联合测量。这些测量的结果将为操控第二个粒子库中的粒子提供必要的信息，根据这些信息人们可以利用第二个粒子库中的粒子复制物体，完成超距传输。

现在你明白了吧，以我们今天的能力，上面讲的每个量子传输步骤都不能完成。一台DeLorean差不多有千亿亿亿个粒子。虽然现在的实验学家们已经能够使多对粒子处于纠缠状态，但离使宏观物体中的粒子处于纠缠还相去甚远。[6]所以，单单准备两间库房，并使其中的粒子处于纠缠就已经远远超越了今天的能力范围。而且，对两个光子

进行联合测量都可以算作是巨大的成功，可想而知，对数以百亿亿计的粒子进行联合测量在今天是超乎想象的。以我们今天的水平估计，客观地讲，要是就用我们传输单个光子时所采用的办法的话，那么实现宏观物体的超距传输将是很遥远的未来——甚至是永远不可能——的事情。

但是，考虑到科学技术领域中最常见的事情就是对不可能的超越，我必须要说，虽然宏观物体的超距传输看起来不可能实现，可是，谁知道未来会怎样呢？40年之前，企业号上的那种计算机看起来也是完全不可能的。[7]

时间旅行之惑

毫无疑问，如果超距传输宏观物体能像打电话叫联邦快递或等地铁那么容易的话，我们的生活肯定会和现在大不一样。那种空想中才有的旅行就会成为现实。旅行或运送的概念将发生革命性的巨变，在便利性和实用性方面的飞跃将重塑人类的世界观。

即使这样，超距传输对我们的宇宙观产生的影响也无法与时间旅行可能带来的冲击相提并论。大家都明白，只要有足够的耐心和决心，我们总能从一个地方到达另一个地方，尽管有的时候只存在理论上的可能性。虽然技术上的问题会使我们的空间旅行受到一定的限制，但只要在满足这些限制的基础上，我们总可以根据自己的愿望决定去哪。但是我们能从现在去到别的什么时刻吗？我们会根据自己的生活经验毫不犹疑地回答：只有一种办法，就是等着那一时刻的到来——等着时钟一秒又一秒地走到那个要去的时刻。显然，我们能去的时刻由不得自己选择，它只能是未来的某一时刻，而不能是过去的时刻。如果要去的时刻比“现在”要早，那么根据经验，我们立即就会得出判断：这绝不可能；回到过去根本不是选项之一。与空间旅行不同，时间旅行毫不理会个人意愿。说到时间，我们只能朝着一个方向不停前进，无论愿意与否。

要是我们能像驾驭空间那样容易的驾驭时间，我们的世界观就不仅仅是发生改变了，那将会是整个人类历史上最了不得的转折点。考虑到这种不可抗拒的强大影响力，我常常会吃惊于竟然没有多少人注意到有一种时间旅行——向着未来去的旅行——的理论基础早在上个世纪早期就已经诞生了。

爱因斯坦一发现狭义相对论时空的性质，就想到了飞向未来的可能性。如果你想看看一千年、一万年甚至是一千万年后地球上的景象的话，你就要看看爱因斯坦的物理定律，它会教你做到这一点。你需要建造一台速度接近光速——比方说达到光速的99.9999999996%——的飞行器。然后你就全速冲进太空，飞上一天，或者10天，甚至27年——这里说的时间指的是你在飞船上感受到的时间。之后你再扭过头来全速返回地球。那么你回来的时候，地球上已经过了一千年、一万年或者是一千万年。狭义相对论的这一预言毫无争议，并且已经在实验上得到了证实；这只是我们在第3章就讲过的速度增加会使时间变慢的一个例子。[8]当然，建造一台速度接近光速的飞行器远远超过了当代的技术水平，所以人们根本没法检验这样一个预言。但正如我们之前讨论过的那样，研究人员已经用其他的一些例子证实了时间变慢的预言，比如说在速度远低于光速的商务飞机上就有速度变慢的情况；又比如说μ子之类的基本粒子以接近光速的速度穿过加速器时也会有时间变慢的现象（静止的μ子会在差不多两百万分之一秒的时间内衰变为其他粒子，也就是说它的寿命大概只有这么长。可μ子运动得越快，它自己的钟就会走得越慢，也就是说它的寿命也就越长）。总之我们可以找到很多使我们相信狭义相对论的正确性的理由；至于使我们怀疑狭义相对论的证据，现在还一条也没有。按狭义相对论的方法飞向未来将全如预言的那样有效。而我们之所以无法进入这样一个时代，完全是因为技术上的不足，而不是理论本身的限制。88

当我们思考另一种时间旅行——回到过去——的可能性时，我们

遇到了一些棘手的问题。当然，你很有可能非常熟悉其中的一些，比如说，回到过去阻碍自己的出生这类标准的时间旅行悖论。在很多科幻小说中，这样的事情都是通过暴力手段达到；但我们也可以采用不这样极端但却同样有效的方式介入过去——比如说阻止你父母的相逢——达到影响你的出生的目的。这里的悖论很清楚：要是你根本没被生出来，那么你怎么可能存在？特别是，你又怎么能够回到过去阻止他们相逢呢？要想回到过去阻止你父母的相逢，你就首先得被生出来呀；但要是你被生出来了，并且回到了过去使你父母不能相逢，那你就不会被生出来了。我们就这样陷入了逻辑上的死循环。

再来看看另一个与此类似的悖论，这个悖论是由牛津的哲学家麦克尔·杜麦特受其同事大卫·多奇启发后提出来的，这个悖论以略有不同、可能更令人困惑的方式捉弄着我们的大脑。现在就来讲一下这个悖论的一个版本。假设我造出了一台时间机器，然后用它到了10年后的未来世界。在“请你吃豆腐”（大规模的疯牛病使人们有了心理阴影。未来的人们早就没了对汉堡包的那种热情。请你吃豆腐一举超过了麦当劳）简单的吃过午餐后，我找到了最近的网吧，想上网看看弦论研究的进展如何。我感到非常惊喜。我发现弦论中所有的未解之谜全部告破。弦论已经被完全研究清楚了，并且已经被成功的用于解释所有已知的粒子性质。人们已经找到了额外维度的确切证据；弦论所预言的超对称粒子——它们的质量、电荷等的性质——已经在大型强子对撞机上全部得到验证。人们再也不需要有所怀疑了：弦论就是宇宙的统一理论。

我进一步探查，看看到底是谁做出了这样重大的贡献，结果我大吃一惊。突破性的论文发表于一年之前，作者是丽塔·格林，我妈妈！我太吃惊了。我并没有不敬的意思：我妈妈人非常好，但她不是一位科学家，也不明白为什么有些人愿意成为科学家。而且，我把我的上一本科普书《宇宙的琴弦》拿给她看，她才翻了几页就扔在一边，因为她一看这些书就头疼。所以，她怎么能写出弦论的那篇关键性论文

呢？我在网上读了她的那篇论文，完全折服于文中简单而深刻的推理。在文章末尾的致谢中，她提到在托尼·罗宾斯研讨会上我让她克服恐惧追寻自己内心深处的物理学家梦想，并对我自那以后多年来在数学和物理方面的细心指导表示感谢。啊！我明白了。原来她也参加了在我启程来往未来之前的那次研讨会。我看我最好回到那个时候再来给些指导。

于是，我又回到了那个时候，开始指导我妈妈学习弦论。可惜进展并不顺利。一年过去了，两年过去了，虽然我妈妈在很努力地学习，可她还是不得要领。我开始变得担忧起来。我们又付出了几年，收效还是甚微。这时，我真的担忧起来了。离她的论文预计发表时间已所剩无几。她到底是怎样写出那篇论文的？最后，我做了一个重大决定。当我在未来读到她的文章的时候，我留下了非常深刻的印象，始终都没有忘记。所以，与其让她自己做出发现——看似越来越不可能了——还不如我来告诉她究竟该怎么写，这样才能确保把我记忆中的那篇论文的每一点都囊括进去。她发表了那篇论文。很快，整个物理世界沸腾起来了。我在我的未来世界读到的每一件事都发生了。

但是这里有一个问题。到底谁应该获得我妈妈的那篇奠基性论文所带来的荣誉？这个人显然不能是我。因为我是通过读她的论文才学到那些理论的。可又怎么能是我妈妈呢？她所写出来的一切都是我告诉她的呀。当然，这里问题的关键并不是谁应该获得荣誉，而是新的知识、新的见解、发表在我妈妈的论文中的新思想到底是哪来的？我到底能指着谁说“就是这个人或这台计算机得出了新的结论”呢？我没有这种洞察力，我妈妈也没有，而且在这里也没有其他人什么事，我们甚至连计算机都没用过。但是，那些精彩的理论不知如何全都出现在她的论文中。很明显，在一个允许既向过去又向未来的时间旅行的世界里，知识会凭空产生。或许这个问题不像阻止你出生那个问题那么令人费解，不过也够古怪的了。

我们到底该如何对待这些悖论和古怪的事呢？我们是否能够得出

这样的结论：朝向未来的时间旅行是由物理定律保证的，朝向过去的时间旅行必须得丢掉？有些人可能会这么认为。但是，我们即将看到，我们总归有办法应对这些讨厌的问题。但这并不意味着我们有可能通过时间旅行回到过去——那是我们很快就要讨论到的另一个问题——而是意味着朝向过去的时间旅行至少不会被我们刚刚讨论过的这些问题排除掉。

反思谜题

回想第5章，我们曾从经典物理的角度出发讨论过时间的流动，并得到了一幅与直观印象全然不同的物体图像。一路谨慎地探求使我们形成这样的看法：时空就像一块冰，时空中的每一个时刻都永远地冻结在那里。这与人们通常的看法——时间就像河流一样，带着我们从一个时刻流向下一个时刻——截然相反。这些冻结的时刻按不同运动状态中的不同观测者的不同方式汇聚成了现在，汇聚成了同一时刻发生的事件。时空块可以被切成不同的“现在”的概念，为了更好地体会这一点，我们也将时空比喻成一块可以从不同的角度切片的面包条。

但要是先把各种比喻放在一边，第5章告诉我们的是：时刻——组成每一块时空条的事件——就是时刻。时刻是永恒的。所有的时刻——每一个事件——都将永远存在，就像空间中的每一个点都永远存在一样。时刻并不是一被观测者的“聚光灯”照亮就活灵活现起来；那样的物理图像虽然符合我们的直觉，但却经不住逻辑分析。真实的情况是，时刻一旦被照亮，便会永远都被照亮。时刻不会改变。时刻永远都是那样。被照亮只是组成时刻的多个不会改变的性质中的一个。从图5.1中，我们可以很清楚地看出这一点。在图5.1中，我们可以看到组成宇宙历史的所有事件；它们全都静止不变地待在那里。关于同一时刻到底发生了什么事件，不同的观测者有不同的结论——因为观测者们从不同的角度切削时空片——但是所有的时间片及其组成事件却是普适的。

量子力学对时间的经典物理图像做了一定的修改。比如说，我们在第12章中曾经看到，在极小的尺度上，空间和时空将不可避免地变得崎岖起伏。但是（参见第7章），要想完全利用量子力学来讨论有关时间的问题必须首先解决量子力学的测量问题。解决之道有很多，其中之一——所谓的多世界诠释——特别适合用来讨论由时间旅行引出的矛盾，我们会在下一节中进行有关内容的讨论。在本节中，我们先停留在经典物理的层次，用时空的冰块、面包条比喻来讨论有关问题。

假设你已经成功地回到了过去，并且阻止了你父母的邂逅。直觉上好像我们都知道那意味着什么：在你回到过去之前，你的父母相遇了——比如说，在1965年12月31日午夜[89]钟响的时候，他们相逢于纽约的晚会——然后又及时地生下了你；又过了许多年，你决定回到过去——1965年12月31日，也就是在这个时候，你改变了一切，使你的父母没有办法相逢，于是你妈妈也就没办法生下你。在前面我们曾经提到，对时间更为合理的描述是所谓的“时空片描述”，现在我们就用时空片描述来看看前面的“直觉式描述”到底有哪些问题。

本质上讲，直觉式描述之所以不正确就是因为假定了时刻可以变化。根据直觉式描述，1965年12月31日午夜钟响的时刻（采用了标准的世俗时间片）“起初”是你父母相逢的时刻，“后来”，由于你的介入，1965年12月31日午夜钟响的那一刻，你父母相隔几英里，甚至根本不在同一个大陆上。这种叙述事件的方式的问题在于把时刻看成是可以变化的了，但我们已经知道，时刻是不变的，它们就只能是本来的那个样子。时空片应当是一种稳固不变的存在。一个时刻根本不可能“起初”一个样，“后来”又一个样。

如果你真的用时间机器回到了1965年12月31日，那你就会出现在那里，你过去一直在那里，将来也一直在那里，你永远都不可能不在那里。不会有两次1965年12月31日——一次你不在那里，一次你又在那里。在图5.1中，你静止不变的存在于时空片中各种不同位置

处。如果你今天决定乘坐时间机器回到1965年12月31日晚间11点50分，那么这个时刻就会出现在那些能找到你的时空片中。你于1965年新年前夜在纽约的露面将是时空永恒不变的性质。

这样一种认识仍会使我们得到一些离奇的结论，但却不会再有悖论了。比如说，你会出现在1965年12月31日晚间11点50分的时空片中，但是在那之前的时空片中完全没有你的任何记录。这虽然非常奇怪，但却并没带来任何悖论。如果有个家伙看到你在11点50分突然出现，他可能会吓坏了，但要是他问你是哪来的，你就可以非常酷的回答："未来。"至少到目前为止，我们还没在这样的场景中发现任何逻辑上的矛盾。当然，更加值得关注的是：你开始执行任务，阻止你父母相遇，这时又会发生些什么呢？要是非得坚持"时空块"观点的话，那么我们就只能这样回答：你肯定无法成功。不管你在新年前夜做些什么，你都终将失败。想要阻止你父母相逢——虽然看似并不难——在逻辑上是绝对行不通的。你父母肯定会在午夜时分相逢。你不会消失，你在那里，你"一直"都将在那里。每个时刻都存在，全都不会发生改变。将变化这一概念用在时刻身上绝对是对牛弹琴。你父母在1965年12月31日午夜钟声响起的时候相逢于纽约，没有任何事情会改变这一事件，因为这一事件是永恒不变的事件，它在时空中永远有一席之地。

事实上，现在你再想想，回忆起童年的时候曾经问过你的父亲他是怎样向你妈妈求婚的，你还记得他说他根本毫无准备。他在求婚前几乎从没见过你妈妈。但就在新年前夕纽约的一个晚会上，他看到一个男子——竟然宣称自己来自未来——不知道从什么地方突然冒了出来，这可把他吓坏了，简直不知所措，以至于一见到你妈妈他就昏头昏脑的决定求婚了。当时就是这样。

关键之处在于时空中全部的不变的事件必须能合成连续的自洽的统一整体。这样的宇宙才有意义。你乘时间机器回到1965年12月31日这件事，根本就是你在履行自己的命运。有一个人突然出现在1965

年12月31日晚间11点50分的时空片中，而在之前的时空片中并没有这个人。要是假想我们身在图5.1之外，我们就会直观地看到这一切；我们也会看到那个出现在1965年12月31日晚间11点50分的时空片中的人就是现在年龄的你。要想使这些几十年前的时空片有意义，你就必须回到1965年。而且，我们这些“世外”之人还看到你父亲在1965年12月31日晚间11点50分战战兢兢地问了你个问题，然后就跑开了，接着在午夜时分遇到了你妈妈；再跳过几张时空片，我们看到你父母结婚了，又过不久你出生了，你慢慢长大了，然后有一天，你进了时间机器。如果真的有可能进行时间旅行的话，我们就再也不可能单单用更早时刻的事件就可以解释某一时刻的事件了（无论从谁的角度看都是如此）。但是，要是把全部的事件放在一起考虑的话，我们就会有一个合理的、连续的、毫无矛盾的故事了。

正如我们在前面一节强调过的那样，无论怎样发挥想象力，这也绝不意味着我们有可能通过时间旅行回到过去。但是上面的分析很清楚的告诉我们，那些所谓的悖论，比如说阻止自己的出生等，根本就是逻辑不清的产物。即使你能乘时间机器回到过去，你也什么都改变不了，就像无论怎样你也不能让π不是3.1415926……一样。如果你真的回到了过去，那你就是，并且永远都是过去的一部分，而这个过去与使你旅行到这里来的过去别无二致。

置身于图5.1之外的话，这种解释严谨又条理分明。纵览整个时空片，我们看到的是环环相扣、井然有序的宇宙连字谜。但是，在1965年12月31日的你看来，一切都那么令人困惑。我在前面的分析中说明，无论你费多大的劲，你都不可能用经典物理的办法阻止你父母的邂逅。你可以眼睁睁地看着他们相遇。你甚至可以为他们的邂逅安排机会；当然，你安排的邂逅可以不用像我在前面随随便便讲的那样。你可以一再的回到过去，过去可能有很多个你，每一个你都想阻止你父母的相逢。但要想成功的阻止你父母邂逅就必须改变一些东西，而这些东西会使改变这一概念失去意义。

但是，即使我们有了这些抽象的思考，我们也会忍不住问这样的问题：你为什么成功不了？如果你就在晚间11点50分的晚会现场，看到了你年轻时代的妈妈，到底是什么使你没法把她带走？又或者说，你看到了你年轻时代的父亲，但究竟是什么使你不能——这么说实在大不敬，但还是说了吧——给他来一家伙？难道你没有自由的意志了吗？从这里开始，量子力学就要登场了。

自由意志，多重世界，时间旅行

自由意志，即使不给时间旅行增加新的麻烦，也是个很棘手的问题。经典物理学定律带有确定性。正如我们在前面的章节中看到的那样，如果你既能准确的知道现在这个时刻事物是怎样的，又能洞悉全部经典物理学定律的话，那你就能准确地说出在任意给定时刻——无论是过去的某一时刻还是未来的某一时刻——事物是怎样的。这样的方程与假定存在的个人的自由意志毫无关系。所以，有些人就会根据这一点得出这样的结论：在经典宇宙中，自由不过是假象而已。你是由粒子组成的，如果物理定律可以确定任意时刻你的粒子的一切——这些粒子都在哪，它们都是如何运动的，等等——那你决定自己行动的意志力就完全可以推算出来。我相信这样的说法，但是有些人——他们认为人类不只是粒子的总和——并不以为然。

无论怎样，这些想法都没什么实际价值，因为我们所生活在其中的这个宇宙遵循的是量子物理，而不是经典物理。在量子世界中，真实世界的物理，虽然跟经典物理所描述的世界有些相似之处，但是还有些重要的区别。正如你在第7章中学到的那样，如果你知道了此时此刻整个宇宙中所有粒子的量子波函数，那么薛定谔方程就会告诉你你所感兴趣的任意时刻的波函数。量子物理在这一点上的确具有经典物理的那种确定性。但是，观测会给量子力学的故事增添很多新的篇章，正如我们所见，有关量子测量问题的热烈争论仍然未熄。如果有那么一天，物理学家们最终认定薛定谔方程就是量子力学的一切，那么整个量子物理，就会具有同经典物理一模一样的确定性。有了经典

物理的确定性，有些人就可以说自由意志不过是假象；而有些人会不同意。但是如果现在的我们错过了量子力学的部分故事——如果从概率性到确定性的结果还需要某些超出标准量子体系的东西——那么自由意志至少有可能在物理定律中找到一个具体的实现。可能真像某些物理学家推断的那样，在未来的某一天，人类会发现有意识的观测行为是量子力学不可或缺的一个元素，是从量子迷雾中提取结果的催化剂。[9]就我个人而言，我觉得这极其不可能，但是我找不到证明它不对的办法。

现在的局面是自由意识的地位及其在基本物理定律中扮演的角色仍不清楚。所以我们接下来就分别分析一下这两种情况——自由意识本就是一种假象；自由意识其实是真实的。

如果自由意识本是假象，并且我们有可能通过时间旅行回到过去的话，那么你无法阻止你父母相逢这件事就没什么好奇怪的了。尽管你感觉你可以掌控自己的行为，但其实不能，真正在背后起作用的是物理定律。当你想过去把你妈妈带走，或者给你父亲来一枪的时候，物理定律就会起阻碍作用。时间机器可能会在错误的地点着陆，使得你到达的时候他们都已经相会了；也有可能在你扣动扳机的时候，枪却卡住了；还有可能你虽然射出了子弹，但却打歪了，直接把你父亲的情敌打中了，反而为你父亲扫清了障碍；再还有这样的可能：当你走出时间机器的时候，你再也没有阻止他们相逢的想法了。不管在你走进时间机器时你在想些什么，你走出时间机器时的行为只能是和谐一致的时空故事中的一部分。物理定律绝不允许任何有违逻辑的行为。你做的一切都会很好地符合逻辑，现在如此以后也将一直如此。你改变不了那些不可改变的事实。

如果自由意识不是假象，并且我们有可能通过时间旅行回到过去的话，量子物理就会给出与经典物理完全不同的说法。另一种大卫·多奇所倡导的非常吸引人的说法使用了量子力学的多世界诠释。还记得我们在第7章中提到过的多世界诠释吗？在这个理论框架下，包含在

波函数中的每一种可能结果——粒子各种可能的自旋指向或者粒子可能出现的位置——都会分别出现在各自的平行宇宙中。我们所看到的任意给定时刻的宇宙都只是量子力学所允许的无限种可能演化中的一种。这个理论体系引人注意的一点在于它所提出的解释：我们觉得自己可以自由的选择做这做那，而这一点反映的是我们在接下来的时刻进入或这或那的平行宇宙的可能性。当然，既然在平行宇宙中你我都有无数个拷贝，那么我们就有必要在这个宽泛的框架下解释一下个人身份与意志的概念了。

就时间旅行和可能的悖论，多世界诠释提出了一种新颖的解决办法。你返回1965年12月31日晚上11：50，拿出你的枪，瞄准你的父亲，扣动扳机。枪响了，你击中了目标。但是这件事情并没有发生在你登上时间机器的那个平行宇宙，所以你的旅行并不是一趟时间旅行，而是从一个平行宇宙来到另一个平行宇宙的旅行。在你扣动扳机击中目标的这个平行宇宙中，你的父母没法相遇——在这样一个宇宙中，多世界诠释保证了我们的存在（因为符合量子力学的所有可能的宇宙都会存在）。所以，根据这样的说法就没什么悖论存在了，因为每个给定时刻都会有位于不同平行宇宙中的各种不同的版本；根据多世界诠释，时空片会有无穷多个，而不是唯一的一个。在原来的宇宙中，你的父母相逢于1965年12月31日，后来你出生了，长大了，不知怎地就和你父亲的关系变得恶劣了，你着迷于时间旅行，开始了你去往1965年12月31日的时间之旅。而在你所到达的那个宇宙中，你父亲被刺于1965年12月31日晚上，那个时候他还没有遇到你妈妈，刺杀他的凶手宣称自己是他未来的儿子。你在这个宇宙中的版本永远都不会被生出来，不过没关系，那个扣动扳机的你的确是有父母的。只不过他们生活在另一个平行宇宙中。至于这个宇宙中的人们是相信你的故事还是把你看成妄想症患者，那我就不知道了。但可以肯定的是，在每一个宇宙中——你离开的那个宇宙以及你来到的这个宇宙——人们都会讨厌自相矛盾的事情。

而且，即使按照现在这种宽泛的说法，你的时间旅行也不会改变过去。对于你离开的那个宇宙，这一点很明显，因为你回到的根本就不是它的过去。至于你去往的那个宇宙，你在1965年12月31日晚上11：50出现也没有改变那个时刻：因为你的确、并且永远都会在那个时刻出现在那个宇宙。于是我们再一次看到，根据多世界诠释，在每一个平行宇宙中发生的事件都是物理上自洽的事件。在你到达的那个宇宙中，的确根据你的心意发生了谋杀事件。你在1965年12月31日晚上11：50的露面，以及你所造下的一切罪孽，都是那个宇宙永远不可抹去的真实之一。

多世界诠释也会对那些不知道从哪突然冒出来的知识——比如我妈妈写出来的那篇弦论方面的发轫之作——给以类似的解释。按照多世界诠释的说法，在无数平行宇宙中的一个中，我妈妈的确迅速的成了弦论方面的专家，我所看到的那篇弦论论文全都是她一个人写出来的。当我决定飞往未来的时候，我的时间机器恰好把我带到了那个她写出了论文的宇宙。我在我妈妈的那篇论文中读到的结果事实上是她在那个宇宙中的版本首先发现的。然后我回到了我的时代，我妈妈在这个平行宇宙中的版本实在是没办法理解物理学了。在多年教而未果的情况下，我选择了放弃，然后告诉她怎样写出那篇论文来。于是我们看到，在这个解释里丝毫没有“理论的关键性突破究竟是谁做出来的”这种困惑。真正的理论发现者是我妈妈在另一个宇宙中的版本，在那个宇宙中的“她”是一个物理学天才。而我的几次时间旅行带来的结果就是将那个宇宙中的我妈妈做出来的发现带给了另一个平行宇宙中的我妈妈。如果你觉得平行宇宙这种解释要比找不到文章作者这件事——极具争议性的命题——好理解的话，那你就为知识和时间旅行的相互影响找到了一个稍好些的解释。

我们在这节或前一节中讨论过的这几种方案都不一定是时间旅行所带来的悖论的真正解释。这些解释方案告诉我们的是，回到过去的时间旅行并不会因为这样的悖论就被排除掉。即使以我们现在的理解

力，物理学家们也能找到很多方法来规避掉这些悖论。但是，没有排除掉远不等于承认它的可行性。所以，我们现在来问几个主要的问题。

时间旅行能回到过去吗

大部分清醒的物理学家都会回答不可能。我也会说不可能。但是这个不可能并不同于有些问题的不可能，比如说，狭义相对论会允许一个有质量的物体达到并且超过光速吗？又或者麦克斯韦理论会允许带一个电荷的粒子分解成总共带两个电荷的几个粒子吗？对于这些问题，我们会毫不犹疑地回答你不可能，这种不可能是经过实践检验的。

事实上，没人能够证明物理定律已经将回到过去的时间旅行排除掉了。正相反，倒是有几个物理学家列出了几条建造时间机器的理论方法。要是你拥有无限的技术能力，可以任意的运用已知的物理学定律，那么你就有可能建造出一台时间机器（我们所说的时间机器指的是既能飞往未来也能回到过去的时间机器）。我们这里说到的时间机器可不是H.G.威尔斯所描述的那个旋转式的小发明或布朗博士那台加大马力的DeLorean[90]。而且所有的设计元素都刚好和已知的物理定律擦边，这就使很多研究人员猜想，等到我们对大自然规律的认识再进步一点，现有的和未来的时间机器方案可能就不仅仅是纸上谈兵了。但就今天而言，这种猜测仅仅基于一种感觉和一些间接证据，没有实际的根据。

至于爱因斯坦本人，在正式发表广义相对论之前的那十几年高强度研究过程中，他也曾经思考过通过时间旅行回到过去的可能性。[10] 坦白地讲，他要是没想过这些可能性那才奇怪呢。因为他对空间和时间激进的诠释摒弃了长久以来的教条，人们自然会不断地问：这场巨变还能持续多久？人们熟知的、来自于日常经验的、纯直觉式的时间性质中还有哪些能够幸存下来？但爱因斯坦之所以没在时间旅行这个

问题上发表什么文章是因为单凭他自己并没能做出什么有价值的东西。但在他的广义相对论的文章发表后的10多年间，另一些物理学家慢慢地做出了点东西。

广义相对论的早期论文中有几篇与时间机器有关，包括：苏格兰物理学家W.J.范·斯托克姆[11]写于1937年的一篇论文和爱因斯坦在高等研究中心的同事科特·哥德尔于1949年所著的另一篇论文。范·斯托克姆研究了广义相对论中的一个假象问题——一个高密度的无限长圆柱体绕着它自身的无限长轴做旋转运动的有关问题。范·斯托克姆对这个并非物理上真实的无限长的圆柱体进行了一番研究，得出了一些有意思的东西。我们在第14章中曾经看到，有质量的旋转体将曳引空间做漩涡状的旋动。在范·斯托克姆所考虑的情况中，这种旋动如此巨大，以至于我们可以根据数学上的分析发现，被拉进漩涡之中的并不只有空间，还有时间。简单地讲，旋动使得圆柱体周围的时间方向发生扭曲，所以绕着圆柱体的圆周运动可以把你带回过去。如果你的火箭绕着圆柱体飞行，那么你就能够返回到你开始这趟旅行之前的那个时刻。当然，没人能够造出一个无限长的圆柱体。但是这篇论文作为早期的一篇暗示广义相对论并不会排除回到过去的时间旅行的论文其价值不可磨灭。

哥德尔在他的论文中讨论的也是一种与旋转物体有关的情况。但哥德尔所关注的并不是在空间中旋转的物体，他感兴趣的是在整个空间都在旋转的情况下会发生些什么。马赫可能会认为这毫无意义。如果整个宇宙都在旋转，那么我们就找不到一个旋转运动可以参照的物体。所以按照马赫的观点，一个旋转的宇宙与一个静止的宇宙毫无区别。而这正是马赫的空间关系概念中另一个没能被广义相对论肯定的例子。根据广义相对论，谈论整个宇宙的旋转是有意义的，而且这种可能性会带来简单的可观测结果。比如说，如果你在旋转的宇宙中射出一束激光，那么广义相对论就会告诉你这束激光将按弯曲路径而不是直线路径射出（你要是骑在旋转木马上，然后用手中的玩具枪向上

射出子弹，那么这颗子弹的路径就有点像我们所讨论的这束激光的路径）。哥德尔的分析之所以令人惊奇在于他认识到：如果你的火箭在旋转的宇宙中按照某一适当的轨迹运行，那你就有可能在你出发之前回到火箭升空之处。因而，旋转的宇宙本身就是一台时间机器。

爱因斯坦对哥德尔的发现表示了祝贺，但也同时提出进一步的探索可能会表明那些使得广义相对论方程允许时间旅行到过去的解会与其他基本的物理要求相矛盾，从而使其失去物理上的意义，变成纯粹的数学意外。至于哥德尔的解，近来精确的实验观测已将其直接现实意义减小到最低的程度，因为观测结果表明我们的宇宙并没处于旋转之中。但是范·斯托克姆和哥德尔的工作放出了瓶中的妖精；之后的几十年间，人们发现了更多使爱因斯坦方程允许向过去时间旅行的解。

近些年来，人们对设计假想中的时间机器的兴趣又浓厚起来了。20世纪70年代，弗兰克·蒂普勒重新分析并且精炼了范·斯托克姆所提出来的解决方案，而1991年，普林斯顿大学的理查德·哥特发现了另一种建造时间机器的方法，他的方法利用了所谓的宇宙弦（假想中早期宇宙所残留下来的无限长、灯丝似的东西）。这些贡献都非常重要，但最容易讲清楚的是由基普·索恩及其在加州理工学院的学生们提出来的一种方案——有关的概念我们曾在早前的章节中有所讨论。索恩他们利用的是蠕虫洞。

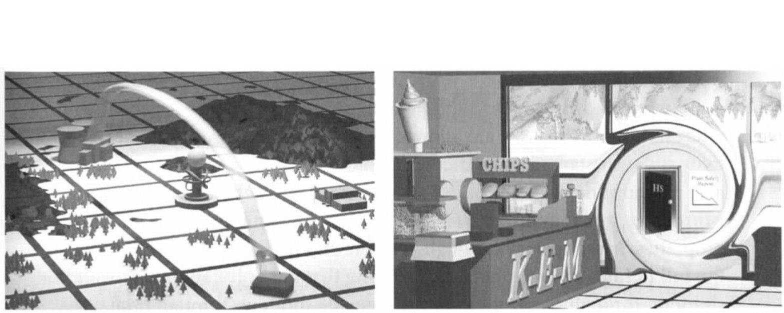
蠕虫洞时间机器的蓝图

我在这里先把建造一台索恩所构想的蠕虫洞时间机器的基本步骤列在这里，在下一节中我们将详细讨论索恩雇佣的承建商将会遇到哪些问题。

蠕虫洞是假想的空间通道。蠕虫洞类似于穿山而过的隧道，那种穿山而过的隧道可以缩短两个不同位置之间的距离（没有它的话你就只好翻过山才能到另一个位置）。蠕虫洞所起的作用就类似于此，但

它同传统意义上的隧道有一个重要的区别。传统意义上的隧道是在已经存在的山体上凿出来的——没有隧道的时候山和山所盘踞的地方就已经在那里了——而蠕虫洞则是沿着一条全新的之前并不存在的空间管构建出一条连接空间中两个不同位置的通道。即使你把穿山而过的隧道废掉，那里的空间仍然存在。但你要把蠕虫洞移除的话，它原本所占据的空间就消失了。

图15.2（a）中所示意的就是连接Kwik-E-Mart和斯普林菲尔德核电站[91]的蠕虫洞，当然这种画法有些误导之处，因为蠕虫洞画成了横跨斯普林菲尔德小镇的样子。更准确一点的话，蠕虫洞应该被想成是一个全新的空间区域，并且这个区域只在其端点——或者说端口——才与我们熟悉的普通空间相交。你要是以为只在斯普林菲尔德大街上走走，抬头看看天就能找到蠕虫洞的话，那你就大错特错了。看到这个蠕虫洞的唯一方法是走进Kwik-E-Mart，在店里你会发现平常的空间中开着一个洞——一个蠕虫洞的端口。往洞里看去，你会看到核电站的内部，也就是图15.2（b）中的第二个端口的位置。图15.2（a）中另一个容易被误解之处在于蠕虫洞看起来似乎并不是一条捷径。将图15.2稍稍改动一下变成图15.3的样子，这个问题就解决了。如你所见，通过常规的路径从Kwik-E-Mart到核电站的确是比通过蠕虫洞要远些。图15.3之所以会画得扭曲完全是因为我们很难在平面上画出广义相对论几何，但即使画成这样，这张图也会给我们带来一些蠕虫洞的直观感受。




（a）　（b）

图15.2　（a）：KWik-E-Mart和核电站之间的蠕虫洞。（b）：视线穿过蠕虫洞看到的景象，洞的一端是KWik-E-Mart，而另一端就是核电站。

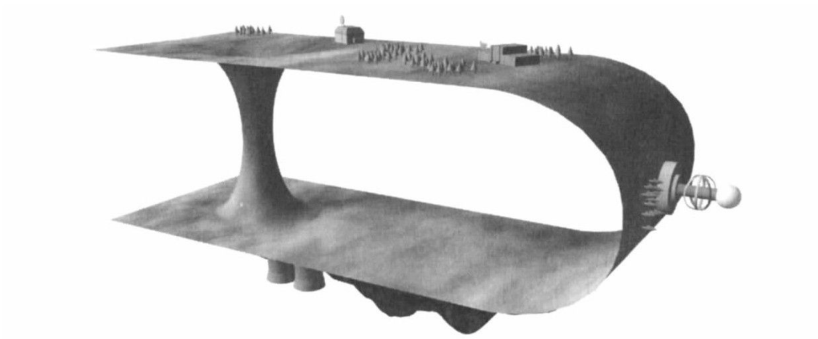


图15.3　利用几何，我们可以更加清楚地看到蠕虫洞真是一条捷径（蠕虫洞实际的端口分别在KWik-E-Mart和核电站内，图中很难显示这一点）。

没有人知道蠕虫洞是否真的存在，但是早在很多年前物理学家们就证明了蠕虫洞是广义相对论的数学所允许的内容，当做游戏搞搞相关理论研究是完全没有问题的。20世纪50年代，约翰·惠勒及其合作者成为蠕虫洞研究方面的先锋人物，他们一道发现了蠕虫洞的很多基本数学性质。到了近一些的时候，索恩及其同事发现蠕虫洞不仅仅是空间上的捷径，也可以是时间上的捷径，从而使蠕虫洞研究领域空前繁荣起来。

现在我们来看看索恩的想法。想象这样一幅场景：巴特和莉莎分别站在斯普林菲尔德蠕虫洞的两端——巴特在Kwik-E-Mart这端而莉莎在核电站那端——懒散地聊着给霍默准备什么样的生日礼物，巴特准备来一趟跨星系的短程旅行（给霍默带点他喜欢的仙女座炸鱼条回来）。莉莎本来对这趟行程不感兴趣，但是考虑到她自己一直很想看看仙女座，所以她就劝说巴特把他那边的端口挂到飞船上，这样她就可以看看仙女座的样子。你或许会认为要是巴特把端口挂到飞船上一起走的话，蠕虫洞就会被抻长，但这是在假定蠕虫洞是普通空间才会有的情况。其实不然。如图15.4所示，由于奇妙的广义相对论几何，蠕虫洞的长度会在整个行程中保持不变。这一点非常关键。即使巴特

的飞船飞往仙女座，他与莉莎之间的蠕虫洞的长度也不会发生变化。而这就使蠕虫洞作为捷径的角色更加明显了。

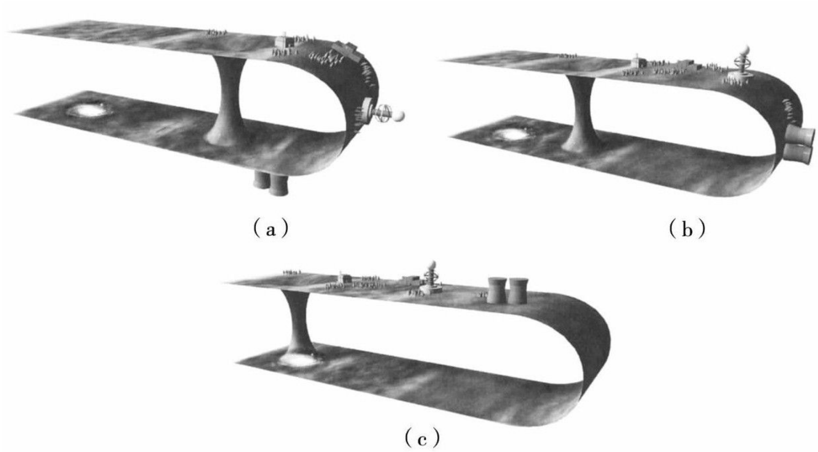


图15.4　（a）：连接KWik-E-Mart和核电站的蠕虫洞。（b）：蠕虫洞开启的下端连的是外层空间（从核电站连到未画在图中的宇宙飞船）。（C）：蠕虫洞的一端到达了仙女座星系，另一个端口则还在KWik-E-Mart。在整个航行过程中蠕虫洞的长度并未发生变化。

为明确起见，我们姑且假定巴特的飞船以99.9999999999999999%光速的速度飞行了4个小时才到达仙女座，在这个过程中他与莉莎一直通过蠕虫洞聊天，就像起飞之前那样。当飞船到达仙女座的时候，莉莎让巴特停下来以便她能清楚的看一看仙女座的风景。但是巴特觉得应该赶快叫上一份外卖炸鱼条，然后掉头返航。虽然莉莎对这么快就返航感觉很气恼，但还是同意继续和巴特聊天。4个小时再加几把五子棋后，巴特安全地降落在斯普林菲尔德高中。

当他从舷窗中望出去的时候，巴特一下子呆住了。那些建筑看起来完全不一样了，滚球场上空漂浮的计分板显示现在已经是他离开之后差不多600万年的时候了。“嘿，哥们，怎么回事！？”他对自己说道，过了一会，一切都搞清楚了。他记起最近通过心连心的方法从杂耍鲍勃那里学到的狭义相对论：你运动得越快，你的钟就会变得越

慢。如果你乘坐高速宇宙飞船出去遛一圈然后回来，那么你飞船上的时间可能仅仅过了几个小时的光景，但是外面的世界已经过了成千上万年。快速的计算后巴特确认，以他航行的速度来看，飞船上的8小时意味着外面世界的600万年。计分板上的时间是正确的；巴特认识到他已经来到了未来的地球。

“……巴特！你在哪？巴特！”莉莎从蠕虫洞里喊道，“你能听到我说话吗？加快油门，我想按时回家吃饭。”巴特看了看蠕虫洞的端口，告诉莉莎他已经降落在斯普林菲尔德高中的草坪上了。从蠕虫洞中仔细地看了看后，莉莎发现巴特没说谎，但从Kwik-E-Mart往斯普林菲尔德高中的方向望过去，莉莎没能发现飞船的影子。“我糊涂了。”她说。

“实际上，这很好理解。”巴特得意洋洋地回答道，“我的确是降落在了斯普林菲尔德高中，只不过是600万年后的斯普林菲尔德高中。你从Kwik-E-Mart的窗户看不到我，因为你虽然看对了地方，但却没看对时间，你早了600万年。”

“对了，没错，狭义相对论的时间延迟效应。”莉莎赞同道，“很酷。但不管怎么说，我想按时回家吃饭，所以快爬过蠕虫洞，我们得快点。”“好的。”巴特应道，然后爬出了蠕虫洞。他在阿普[92]那儿买了黄油棒，然后就和莉莎回家了。

注意，虽然巴特一会儿就穿过了蠕虫洞，但这却意味着他穿过了整整600万年。巴特与他的飞船还有蠕虫洞的端口降落在600万年后的未来。他只需要走出飞船，同人们谈上几句，看看报纸，就会确认这一切都是真的。但是，一旦他爬出蠕虫洞，就会又回到现在和莉莎在一起了。任何一个像巴特一样穿过蠕虫洞的人也都有同样的遭遇：他也将穿越600万年的时间。与此类似，任何人从位于Kwik-E-Mart的蠕虫洞端口爬进去都会来到巴特飞船降落的600万年后。这里的关键之处在于巴特并不是仅带着蠕虫洞的端口穿越空间。他的这趟旅行也

使蠕虫洞的端口穿越了时间。巴特的旅程把他自己和蠕虫洞的端口带到了未来的地球。简而言之，巴特将一条空间隧道变成了一条时间隧道；他将一个蠕虫洞变成了时间机器。

图15.5是这个过程的一张粗略示意图。在图15.5（a）中，我们可以看到一个蠕虫洞将空间中的两个位置联系起来了，之所以这样画蠕虫洞是为了使之区别与普通空间。在图15.5（b）中，我们展示了这个蠕虫洞的演化，这里假定了蠕虫洞的两个端口全部处于静止状态（时间片为静止观测者的时间片）。在图15.5（c）中，我们展示了当蠕虫洞的一个端口被挂在往返行程的飞船上时的情况。对于运动的蠕虫洞端口来说，它的时间就像运动的钟的时间一样会变慢，所以运动的端口被传送到了未来（如果运动的钟仅仅过了1个小时，而静止的钟却过了千年的话，运动的钟就来到了静止的钟的未来）。因而，静止的蠕虫洞端口连接的并不是同一时间片的端口，它通过蠕虫洞连接的是未来时间片的端口，如图15.5（c）所示。除非蠕虫洞的端口又运动了，否则的话两个端口之间的时间差就会一直存在。任意时刻，只要你从一个端口进去再从另外一个端口出来，你就成了时间旅行者。

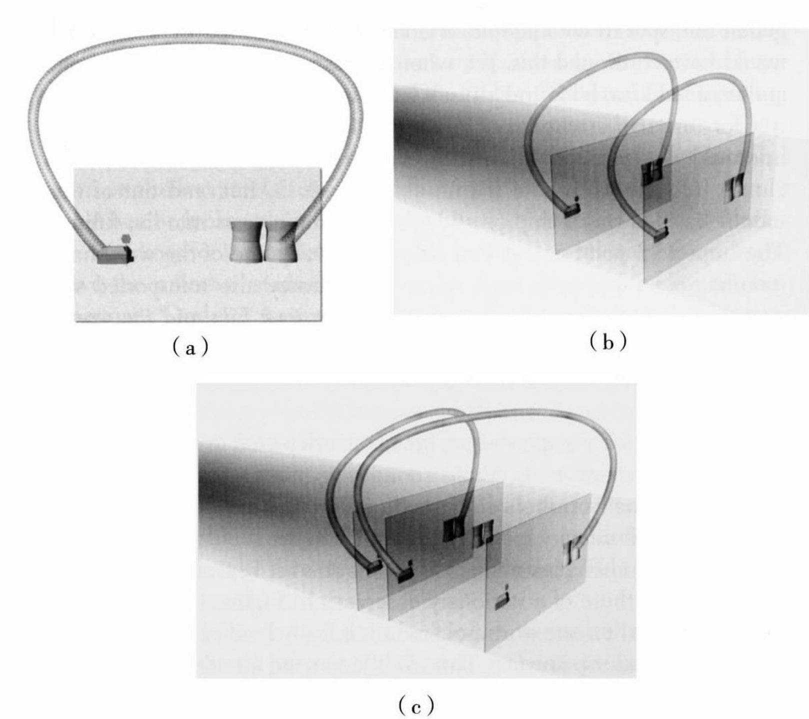



图15.5　（a）：某个时刻所创建的蠕虫洞，将空间中的两个位置联系起来了。（b）：如果蠕虫洞的两个端口之间没有相对运动，它们就会以相同的速率“穿过”时间，这样一条蠕虫洞连接的是相同时间的两个区域。（C）：如果蠕虫洞的一个端口处于往返旅程中（图中未示），那它所用掉的时间就会少些，因而这样的蠕虫洞连接的就是不同时刻的两块空间区域，这样的蠕虫洞就是一台时间机器。

## 建造一台蠕虫洞时间机器

现在，我们已经搞清楚了建造一台时间机器需要有哪几步。第一步：找到或者干脆创造一个蠕虫洞，这个蠕虫洞要宽到能把你或者你想通过时间机器传送的东西放进去的程度。第二步：使得蠕虫洞的两个端口之间存在时间差——也就是说，使一个端口相对于另一个端口运动。理论上讲，这就行了。

但在实践中能行吗？正如我在开始时提到的，人们甚至不知道是

不是真的有蠕虫洞存在。有些物理学家说在空间结构的微观层次上存在很多微小的蠕虫洞，这些小蠕虫洞产生于引力场的量子涨落。如果是这样的话，那么我们面临的挑战就是如何将它们放大到宏观尺度。至于究竟应该怎么达到这种效果人们已经提出了一些建议，不过那些方案都非常晦涩，不在有趣的理论妙想范围内。而另一些物理学家则将制造宏观蠕虫洞当做广义相对论应用方面的工程项目。我们知道空间反映的是物质和能量分布，所以，只要能很好的控制物质和能量，我们就有可能使某块空间区域变成蠕虫洞。这种方法还有另外一个难点，正如我们非得在山上撕开一个口子才能建造穿山隧道一样，我们也必须在空间结构中撕开一个口子才能给蠕虫洞找一个端口。[12]没人知道物理定律是否允许我们这么干。我在弦论方面的工作（参见第13章“膜”小节）表明，某些类型的空间撕裂是没问题的，但我们目前为止还不清楚这样一些缝隙是否与制造蠕虫洞有关。即使按最好的情况来看，获得宏观蠕虫洞的想法也是一个非常长的时期内才有可能实现的梦想。

而且，就算我们想出了什么办法得到了宏观蠕虫洞，事情也没完，我们还要面临一大堆的困难。首先，早在20世纪60年代，惠勒和罗伯特·弗勒就曾利用广义相对论方程证明，蠕虫洞是不稳定的。蠕虫洞壁会一瞬间就向内塌陷，这就使得我们没法用它做任何旅行。但是后来，物理学家们（索恩、莫里斯以及马特·维瑟）又发现了绕过塌陷问题的可能途径。如果蠕虫洞不是空的，而是含有具有外推力的物质——所谓的奇异物质（eXotic matter），那么蠕虫洞就有可能持续开放并且稳定。虽然奇异物质的效应有点类似于宇宙常数，但奇异物质生成外推的排斥性万有引力纯粹是由于具有负能（而不是宇宙常数的负压特征[13]）。在某些非常特别的情况下，量子力学会允许负能量的存在[14]，但生产足够多的奇异物质来维持宏观蠕虫洞的开放显然是一项巨大的挑战（比如说，维瑟曾经计算过，要想打开1米宽的蠕虫洞，所需要的负能差不多等于太阳在过去100亿年间产生的总能量[15]）。

其次，即使我们通过什么方法找到或者制造出了宏观蠕虫洞，并且有办法支撑洞壁使之不塌缩，还使两个端口之间有了一定的时间差（比如说，让其中的一个端口绕另一个端口高速飞行），要想得到一台时间机器也还得克服另外一些困难。包括霍金在内的几位物理学家提出了新的可能性。他们指出，真空涨落——由于各种场的量子不确定性而带来的涨落，如我们在第12章中讨论过的那样，即使对于平坦空间，这种涨落也是存在的——可能会在蠕虫洞变成时间机器的时候摧毁它。其原因在于就在穿越蠕虫洞的时间旅行变得可行时，一种可怕的反馈机制——就像有的时候我们拿的麦克风与音箱的位置不合适而产生刺耳的啸叫——可能开始起作用。未来的真空涨落可能也会穿过蠕虫洞而回到过去，而回到了过去的真空涨落可能也会顺着普通的空间和时间而来到未来，再次进入蠕虫洞，再次回到过去，如此循环往复使蠕虫洞中出现一直增加的能量。推测起来，这样的能量可能会毁灭蠕虫洞。理论研究表明这是一种真正的可能性，但真要实际计算又需要我们理解弯曲时空中的广义相对论和量子力学，因此，这一问题尚未有定论。

显然，建造一台蠕虫洞时间机器将面临巨大的挑战。但是我们还不能下定论究竟是能还是不能，除非我们对量子力学和引力的理解更上一层楼，而这种提升可能需要通过弦论的发展。尽管单靠直觉，大部分物理学家都认为通往过去的时间旅行并不可行，但在今天，这个问题尚未盖棺定论。

宇宙观光

谈到时间旅行的时候，霍金提出了一个有趣的问题。他问道："如果真的可以做时间旅行，那么我们为什么从来也没遇到过来自未来的客人呢？"你可能会说："或许我们早就遇到过了。"你也可能进一步辩解道："这些时间旅行者中的大部分都被我们扔到紧锁的病房里去了，剩下的那些多半不敢表明自己的身份。"当然，霍金的话半是玩笑，我的回答也是这样，但他提出的问题却很严肃。如果你也

像我一样认为，我们尚未遇到过任何来自未来的人，那是不是可以说时间机器根本是不可能的呢？毫无疑问，要是未来的人们真的成功的建造出了时间机器，那么某些历史学家必定要亲身到现场研究一下第一颗原子弹的建造，第一次的月球之旅，或者第一台电视机的投产。所以，要是我们相信从未有来自未来的人访问过，那么也就相当于说我们相信根本不可能有时间机器这回事。

但实际上，这可不是一个必然的结论。乘坐时间机器并不能回到第一台时间机器建造出来以前的时代。对于蠕虫洞时间机器来说，仔细看看图15.5就会明白这一点。尽管蠕虫洞的两个端口之间有时间差，尽管这种时间差使得我们可以向未来或向过去做时间旅行，但你却不能去一个早在时间差建立起来之前的时刻。蠕虫洞本身并不能存在于时间片左边很远的地方，所以你也没办法到达那样的地方。因而，假如说时间机器在距今1万年后建造出来，那一时刻无疑将吸引很多时间旅客，但是在那之间的时代，比如说我们这个时代，就是时间机器永不可及的了。

我们对自然定律的现有理解不仅能告诉我们如何避免显而易见的时间旅行悖论，还能为建造时间机器这样的想法出谋划策，这一点实在令我感到非常新奇。可别当我错了，我可属于清醒物理学家行列——我们凭直觉认定未来的某一天利用时间机器回到过去的可能性将会被物理定律排除掉。但在明确的证明出现之前，我们最好公正客观的保持着开放式思维。至少，研究这些问题的物理学家们不自觉的深化了我们对极端条件下的空间和时间的理解。他们可能正朝着时空高速路迈出关键的第一步。不管怎么说，在我们成功的造出时间机器之前的每一个时刻都将一去不返，永远的弃我们以及我们之后的人类而去。

## 第16章　幻象的未来

空间和时间的前景

对于物理学家来说，其一生的大部分时间都在困惑的状态中度过，这甚至可算是一种职业病。要想在物理学领域获得成功就得学会在通往真理的曲折道路上爱上疑惑。对含混不清的东西无法忍受的感觉激发了普通人身上所蕴藏的非凡天赋和创造性；有待调和的不谐事物特别容易使人集中思想。但是在探索之路上——在为解决著名问题而进行的研究工作中——理论学家们必须穿越迷茫的丛林，他们所能依靠的导引只能是直觉、模糊的概念、断断续续的线索与计算。而且，由于大多数的研究者有掩饰痕迹的习惯，所以在很多复杂困难的细节上人们常常一无所知。但是千万不要忘记没有什么可以轻易获得。大自然不会轻易说出她的秘密。

在本书前面的很多章节中，我们读到了很多有关人类追寻空间和时间之意义的故事。尽管我们已经谈到了很多既深刻又令人惊奇的思想，但是仍有种种疑惑未能弄清楚，我们仍未到达可直呼“Eureka”[93]的那一刻。毫无疑问，我们仍在丛林中探寻出路。那么，未来的路在何方？时空故事的下一章应该是些什么内容？显然，没有人知道。但是近年来，一些线索已初见端倪；虽然这些散碎的证据还不能连成完整一致的物理图像，但很多物理学家仍然相信这些证据暗示着宇宙学上另一次大突破的来临。于此之际，很多物理学家相信时间和空间这样的概念不过是埋藏在物理真实性中的更加精细丰富的基本理论的一种幻象。在即将结束本书之际，让我们试着了解一下这些零散的线索并初窥我们探寻宇宙结构之旅的下一站在哪个方向。

空间和时间是基本概念吗

德国哲学家伊曼努尔·康德认为，要是把空间和时间的概念剔除掉，思考并且描述宇宙就不单单是有难度了，而是根本就不可能。坦率地讲，我能理解康德为什么会这么想。无论什么时候，只要我坐下来闭上双眼，试着想一想某个或某件既不占任何空间也没经历任何时间的东西或事情，我都会大脑短路。有空间才会有内容，有时间才有变化，两者总是无声无息地存在着。具有讽刺意味的是，埋头于数学

计算（这时候用的常常是时空）的时候，才是我的大脑与时空的概念联系最紧密的时候，因为这些数学练习能够短暂的吞没我的思想，在某种抽象意义下就像没有空间和时间一样。但是思想本身以及思想需要依托的身体则还是不得不占据着时间和空间。要真的清除了空间和时间，那么甩掉你的影子简直易如反掌。

然而，很多当代主流物理学家怀疑，空间和时间虽然无处不在，但却不一定是真正的基本概念。就像炮弹的坚硬源于组成炮弹的原子的集群性质，玫瑰的香气源于组成玫瑰的分子的集群性质，猎豹的敏捷源于其肌肉、神经以及骨骼的集群性质，空间和时间的性质——这本书中所讨论的主要东西——也可能源于某种我们尚不清楚的基本成分的集群性质。

物理学家有时会将这些可能性归结为一句话：时空根本是一种幻象——一种有争议的描述，其意义需要合理的解释。不管怎么说，当你被炮弹击中，或者闻到了玫瑰的香气，或者正好看到了飞奔的猎豹的时候，你是无论如何也不会因为这些事物都有其更加基本的结构而否定其存在的。相反的，我想我们大多数人都会同意这些由各种物质汇聚起来的事物自有其存在性，而且，要想了解这些事物，我们仅仅知道作为其组分的原子或分子的性质是不够的，我们还需要研究其整体的性质。但是，由于它们都是由更基本的物质组成，所以我们不会去试着建立一个基于炮弹、玫瑰以及猎豹的宇宙理论。类似的，即使空间和时间真是某种复合实体，那也并非就意味着我们所熟知的空间和时间表象——不论是牛顿的桶还是爱因斯坦的引力——只是一种幻象；无需怀疑，不论未来我们对时间和空间的理解有什么样的发展，空间和时间在实际意义上仍将保有其无处不在的地位。而且，复合时空的概念意味着一种对宇宙更为基本的描述——既不需要空间也不需要时间的宇宙——尚未被发现。这样，空间和时间这种幻象就只是人类所臆造的概念之一，对宇宙最深刻的解读将使我们看清楚空间和时间这一信念将土崩瓦解。当你在原子以及亚原子水平上研究物质的时

候，炮弹的硬度、玫瑰的香气、猎豹的速度都不再有任何意义；与之类似，当我们钻研大自然的最基本法则的时候，空间和时间的概念也将自行消融。

时空并不是最基本的宇宙组成可能使你觉得有点牵强。在这一点上你的感觉可能是对的。但有关时空并不与最基本的物理定律相关联的说法并不是一个胡诌的理论。相反，这样想法的提出正是基于一系列理性的思考。下面我们一起来看看其中最出色的几个想法。

量子平均

在第12章中，我们曾讨论过空间结构以及宇宙中的其他事物究竟是怎样被归结为量子不确定性的涨落的。你或许还记得，正是这些涨落，一下子冲垮了点粒子理论；也正是这些涨落，使得人们无法从点粒子理论中得到一个合理的量子引力理论。弦论则不然，它用圈和片代替了点，抹平了涨落——从根本上减小了量子涨落的幅度。就是这样，弦论成功的统一了量子力学和广义相对论。然而，大大减弱了的时空涨落仍然可以存在（就像图12.2中倒数第二层所展示的放大效果），而我们将以它们为线索来思考时空的命运。

首先，我们了解到，人们所熟知的空间和时间——就是浮现在我们脑海中、运用在方程式中的那个空间和时间——来自于某种平均过程。当你的脸贴近电视机荧屏的时候，你看到的是一个个像素。这时的感觉与你在一个舒适的距离看电视时的感觉大为不同。这是因为，当你的眼睛没法分辨单独的像素时，它就会把各个像素组合起来平均一下，从而得到平滑过渡的图像。请你注意，只有通过这样的平均过程，由大量的像素构成的图像看起来才是连续的画面。同样，时空的微观结构中也存在着随机波动，但是由于不能在那么小的尺度上分辨时空，我们没法直接知晓这些随机波动。我们的肉眼，甚至我们最为强大的设备，会将这些波动平滑起来得到均匀连续的感觉，就像看电视那样。由于这些波动是随机的，因而在一个小区域内“向上”的波动

很可能同“向下”的波动一样多，这样平均一下，大体上就会彼此相消，使我们看到一个平和的时空。但是，与电视机的情形类似，时空之所以以平滑安稳的形式出现，仅仅是由于平均过程的存在。

我们熟悉的时空其实是一种幻象，对于这一说法，量子平均提供了一个非常实际的解释。平均方法的用途很多，但天生就有不能为深层细节提供清楚图像的特点。尽管平均说来，每个美国家庭有2.2个孩子，但你能找出一个有2.2个孩子的家庭吗？尽管一加仑牛奶的平均价格是2.783美元，你却找不到一家按这个价格卖牛奶的商店。我们熟悉的时空也是这样，尽管其自身为平均过程的结果，但却不能描述那些我们想要将其称为“基本”的概念的细节。空间和时间只是一种近似，一种集群概念，一种在除最微观尺度外的几乎所有尺度上研究宇宙的极好工具，一种类似于有2.2个孩子的家庭的幻象。

我们来看看第二种相关的见解：尺度愈小则量子涨落越强意味着不断地将距离或间隔分割成更小单位的概念可能会在普朗克长度（$10^{-33}$厘米）与普朗克时间（$10^{-43}$秒）附近走到尽头。我们在第12章中讨论过这一想法。我们曾经强调过，尽管这一概念与我们从日常生活中获得的空间和时间的感受完全不同，但与日常生活有关的性质被推广到微观尺度时变得面目全非实在没什么值得惊奇的。既然空间和时间的无限可分性是我们日常生活中所熟知的性质，那么这一概念的不再适用就成了另一条线索：暗示我们微观尺度上隐藏着某些我们不了解的东西——或许可称为时空基底的半成品的东西——它构成了我们所熟悉的时空概念。我们认为这尚未明确的组分，这最基本的时空体，不可以再被分成更小的片；因为我们最终会到达量子涨落非常猛烈的尺度，而在那里将没有我们日常在大尺度上感受到的时空。看起来基本层次上的时空组分——不管它到底是什么——被平均过程彻底地改变了面貌，以至于成为我们在日常生活中所体验到的时空。

因而，在最深层次的大自然定律中寻找熟悉的时空可能就像一个音符一个音符的听贝多芬的第九交响曲，或者就像一笔一笔的看莫奈

的画。在最基本的层面上，自然界中的时空就像这些人类表现力中的极品一样，其整体与其部分完全不同，全无类似之处。

## 翻译中的几何

另一种想法，物理学家称之为几何对偶性的理论，同样提出时空可能并不具有基本性，只不过提出这一论断的角度相当不同。相比于量子平均，描述这一理论时需要更多的技术术语，因此你要是觉得这一小节太过晦涩，那就略过不读好了。不过很多研究者都认为这些想法是弦论中最具象征意义的内容，因而努力读一读，试着了解其主旨还是很值得的。

我们在第13章中看到，本以为不同的5种弦论究竟为什么是同一种理论的5种翻译。我们在众多内容中特别强调这一点，是因为很多时候这种翻译会使极其困难的问题变得简单一些。但是，在统一5种理论的翻译字典中，有一种目前为止我一直忽略没提的特性。将一个问题从一种弦论中的形式翻译到另一种弦论中的形式时，其困难度会大幅度变化；同样的，将一个问题从一种时空的几何形式描述转换到另一种时空的几何形式描述时，问题的困难度也会有大幅度变化。我要说的就是这个。

在我们的日常生活中，空间只有三维，时间只有一维，而弦论则需要更多的维度。这就使我们有理由探讨额外的维度究竟藏在哪里这样的问题，就像我们在第12章和第13章中做的那样。我们当时找到的答案是这些额外的维度会蜷曲起来，它们蜷曲以后的尺度如此之小，以至于我们目前的实验水平还无法触及。在那些章节中，我们也弄明白了在我们熟悉的大维度上的物理依赖于额外维度的精确大小以及形状，因为额外维度的几何性质会影响弦的振动模式。很好，现在来谈谈我之前故意忽略的问题。

能够将一种弦论中提出的问题翻译为另一种弦论中的不同问题的

那本字典，也可以用来将第一个理论中的额外维度的几何翻译为第二个理论中的额外维度的几何。比方说，假如你正在研究IIA型弦论的物理内涵，在这个理论中，额外维度蜷曲到特定的大小和形状；那么，你所得到的每个结论至少在理论上可以从被翻译到其他弦论（比如IIB型弦论）中的同一问题中推演出来。但是，要想在不同的理论中完成问题的翻译，就得要求IIB型弦论中的额外维度按特定的几何形式蜷曲，而这特定的几何形式将取决于——但通常来说却区别于——IIA型弦论中的额外维度的几何形式。简而言之，具有按某种几何形式蜷曲的额外维度的给定弦论等价于——可被转译为——具有按不同几何形式蜷曲的额外维度的另一种弦论。

而且，时空几何上的差异并不会很小。比方说，要是IIA型弦论中的一个维度被蜷曲为一个圆环，如图12.7所示，那么那本字典就会告诉你这个理论完全等价于其中一个额外维度蜷曲为圆环的IIB型理论，而且，IIB型弦论中的圆环半径反比于IIA型弦论中的圆环。如果一个圆环很小，另一个就会很大，反之亦然，因而没有任何办法可以分清两种几何（这里我们用普朗克常数的倍数来表示距离。如果一个圆环的半径为R，则数学字典会告诉你另一个半径为1/R）。你可能会觉得搞清楚哪个是大圆环哪个是小圆环非常容易，但在弦论中可真的没有那么容易。所有的观测都得自于弦之间的相互作用；而这两种弦论，具有一个大圆环维度的IIA型弦论和具有一个小圆环维度的IIB型弦论，只不过是同一物理的不同翻译版本——只不过是不同的表述方式而已。你在一种弦论框架下描述的所有观测都可以在另一种弦论中找到完全等价的描述，尽管每种理论的语言以及所给出的解释会所有不同。（之所以有这种可能则是因为，对于沿圆环维度运动的弦来说，可能存在两种定性上全然不同的构造：一种是弦像绕在锡杯上的橡胶圈那样绕在圆环维度上；另一种则是弦被固定在圆环维度的某处而不是绕着它。前者具有正比于圆环半径的能量[半径越大，缠绕它的弦就要被拉伸得越长，因此弦中蕴藏的能量也就越大]，后者具有反比于半径的能量[半径越小，弦在圆环维度所能占据的就越多，所以弦

运动时由于量子不确定性而导致的能量也就越大]。值得注意的是，要是我们可以将原本的圆环半径反换，并同时将“蜷曲的”弦同“不蜷曲的”弦交换，则物理能量——以及更普遍意义上的物理——将不受影响。这一点正是字典在将IIA型弦论翻译为IIB型弦论时所需要的，也正是为什么两种显然不同的几何——大圆环维度和小圆环维度——可以彼此等价的原因所在。）

若我们将额外维的形状由简单的圆环变为第12章中讲过的更为复杂的卡拉比丘流形，我们还将得到类似的想法。额外维度蜷曲为特定的卡拉比丘流形的某种弦论可以被字典转译为额外维度蜷曲为不同的卡拉比丘流形的另一种弦论（我们说两者互为镜像或者对偶）。在这些例子中，不仅卡拉比丘流形的大小可以不同，它们的形状，包括其上的洞的种类和数目，也可以不同。但是那本翻译字典却能保证它们按照正确的方式有所区别，因而，尽管额外维度的尺寸和形状不同，不同的理论所蕴含的物理却是完全相同的（在一个给定的卡拉比丘流形中可能有两种类型的洞，弦的振动模式——以及所诱导的物理——仅对两种类型的洞的数目之差敏感。所以，如果在某个卡拉比丘流形上有两个第一种类型的洞，以及5个第二种类型的洞；而在另一个卡拉比丘流形上有5个第一种类型的洞，以及2个第二种类型的洞，那么虽然这两个卡拉比丘流形的几何形状不同，但它们对应的理论将带来同样的物理[94]）。

从另一个角度看，这对“空间并非基本概念”这一猜测是一个有力的支持。用5种弦论中的某一种描述宇宙的人将会宣称包括额外维度在内的空间具有某种特殊的大小和形状；而用另一种不同的弦论描述宇宙的人将会宣称包括额外维度在内的空间具有另一种不同的大小和形状。这两个人观测的是同一个物理宇宙，但却给出了两种不同的数学描述，而这并不意味着两者中必有一个是错的。他们可能都是正确的，尽管他们有关空间的结论——大小和形状——不尽相同。还需要注意的是，这里并不是说他们按照不同的但却等效的方式切割时空，

就像狭义相对论中那样。这两位观测者所不能达成共识的乃是时空自身的整个结构。这就是关键之所在。如果时空真的是基本的，大多数物理学家都会认为每个人，不论其视角如何——不管他们用的理论语言是怎样的——都将会就时空的几何性质达成共识。但现在的情况是——至少在弦论中是——使用不同理论的人们不能就时空结构达成共识，而这一点正意味着时空很可能是个次级现象，而不是基本概念。

因而我们就自然地被带到这样一个问题前：如果前面两节所讨论的线索的确为我们指引了正确的方向，即我们熟悉的时空只不过是某种基本实体的大尺度表象，那么这种实体究竟是什么？它又具有哪些性质呢？今天还没有人能够回答这个问题。但是在探寻这个问题的路上，研究者们发现了更多的线索，而最重要的线索来自于对黑洞的思考。

黑洞的熵有什么用

黑洞可算是宇宙中最难以捉摸、最老谋深算的家伙。从外部来看，黑洞要多简单有多简单。黑洞的3个特性分别是质量（质量将决定黑洞的大小，也就是其中心到其视界——一旦触碰则有去无回的隐蔽表面——的距离）、电荷，以及自旋速度。这就是黑洞。只要搞清了这几点就可以看到黑洞向宇宙展现的面貌。物理学家将这一点总结为“黑洞无毛”，也就是说黑洞缺乏那种有个性的具体特征。如果你见过了一个具有特定质量、电荷和自旋（当然，你只能间接地通过环绕黑洞外的气体和星体来获得这些信息，因为黑洞是黑的）的黑洞，那你就见过了所有具有相同质量、电荷和自旋的黑洞。

然而，在黑洞岩石般坚硬的外表下，却匿藏了现今所知的宇宙中最极端的狂乱。在一切具有给定尺寸的任意可能组成的物理系统中，黑洞包含着最高可能性的熵。根据第6章讲过的内容，我们可以粗略地讲，这一结论得自于熵的定义——熵是物体内部组分重排数目的量度，与物体的外在无关。至于黑洞，虽然我们并不知道在其内部究竟

有些什么——因为我们根本不知道物质撞进黑洞后会发生些什么——我们却可以心安理得地说黑洞内部成分的重新排列组合并不会对黑洞的质量、电荷和自旋产生影响，就像把《战争与和平》的页码打乱重拍不会影响这本书的重量一样。既然质量、电荷以及自旋完全决定了一个黑洞展现给外部世界的面貌，那么所有的那些重拍组合我们可以统统无视，于是我们说黑洞具有最大可能的熵。

即使讲了这么多，你还是可能会提出用下面这样的办法增加黑洞中的熵。首先造一个与给定黑洞同样大小的空心球，然后往里充气（氧气、氢气、二氧化碳，什么都行）并使之遍布整个球内部。你充进去的气越多，气体的熵就会越多，因为气体分子越多意味着可能的重拍组合数目就越多。于是你就会说，只要你不停地充气，这样一直下去，早晚会超过黑洞的熵。这个办法听起来挺高明，但是广义相对论却告诉你这行不通。你充进去的气越多，球体所包容的质量就越大。在你还没来得及将它充到等体积的黑洞所含有的熵的时候，球体内不断增加的质量就会达到一个临界值，这时球体及其内部的一切就会变成一个黑洞。我们其实找不到什么办法绕过这一点。黑洞就是拥有最大的混乱度。

如果你继续往黑洞里打气，其中的熵将会怎样增加呢？熵当然会继续增加，只不过游戏规则却需要变一变。当有物质跃入黑洞贪婪的视界内时，黑洞的熵毫无疑问会增加，但是其大小也会增加。黑洞的大小正比于其质量，因而，只要你往里注入更多的物质，黑洞自身就会变得更重更大。所以，一旦你通过制造黑洞的办法使某一区域内的熵达到了最大值，那就再也找不到什么增加这个区域内的熵的办法了。这一区域内的混乱度再也不能增加了。其中的熵已经满了。无论你做什么，不管你是充进更多的气还是扔进去一辆悍马，你都只能使黑洞所占据的空间增加。因而，黑洞中所包含的熵的多少告诉我们的不仅仅是黑洞的基本性质，还会告诉我们空间自身的某种基本性：某一空间区域——任何空间区域内，不分地点、时间——内所能容纳的

最大的熵等于与该区域等大小的黑洞中所能容纳的熵。

那么，给定尺寸的黑洞中到底能够包含多少熵呢？这正是有趣之处。我们先靠直觉猜想，用比较容易想象出来的事物来举例，比方说我们先来分析一下特百惠[95]塑料容器中的空气的熵。如果你把两个一样的特百惠容器对接在一起，那么体积就会翻倍，相应的空气分子数目也会翻倍，于是你就会认为熵也翻倍了。具体的计算[1]也将确证这种猜测，并会同时告诉你其他一切（比方说温度、密度等）没有发生变化。我们熟悉的物理系统中的熵正比于其体积。于是我们自然会猜测那些我们不太熟悉的物理系统中的熵也将正比于体积，因此我们就会得出结论说黑洞中的熵正比于其体积。

20世纪70年代，雅克布·贝肯斯坦与史蒂芬·霍金发现这种猜测并不正确。他们的数学分析证明黑洞的熵并非正比于其体积，而是正比于其视界的面积——粗略说来是正比于其表面积。这是一个全然不同的结果。要是你把黑洞的半径翻倍，其体积就会变为原来的8（$2^3$）倍，而其面积则只变为原来的4（$2^2$）倍；要是你把黑洞的半径变为原来的100倍，其体积就会变为原来的100万（$100^3$）倍，而其面积则只变为原来的1万（$100^2$）倍。黑洞体积的变化速度要大于黑洞表面积的变化速度。[2]因而，虽然黑洞所能够包含的熵是所有给定大小的事物中最多的，但贝肯斯坦与霍金却证明黑洞所包含的熵比我们之前简单估算的结果小得多。

熵正比于表面积并不仅是黑洞同特百惠容器之间的一个古怪差别，要是那样的话我们就可以做个笔记然后继续往后讲。我们已经知道（即使只是在理论上），黑洞为空间区域内所能填充的熵的数量设定了上限：取一个与所要讨论的区域等大小的黑洞，找到这个黑洞中的熵的大小，这个数值就是所要讨论的空间区域中所能容纳的熵的上限。既然这个熵，如贝肯斯坦与霍金证明的那样，正比于黑洞的表面积——黑洞的表面积正好等于所讨论的区域的表面积，因为我们假定两者具有相同的大小——我们就可以得出结论说任意给定空间区域内

所能容纳的最大熵正比于该区域的表面积。[3]

这里所得到的结论同思考塑料容器中的气体所得到的结论（在那段讨论中我们发现容器中的熵正比于容器的体积而不是表面积）之间的分歧很容易解释：因为我们事先假定了容器内的气体分子均匀分布，所以我们在讨论塑料容器时实际上忽略了引力；可是不要忘了，当引力起作用的时候，事物会聚团。当密度很低的时候，忽略引力是没问题的，但如果所讨论的问题牵涉到很大的熵，密度很高，那引力就要起作用了，而那段用特百惠塑料容器所做的分析就不再有效了。这种极端条件需要贝肯斯坦与霍金那种基于引力的计算，得到的结论也应当是空间区域中所能容纳的最大熵正比于其表面积，而不是体积。

话说回来，我们为什么要在乎这些呢？原因有两个。

首先，熵界提供了另外一条超小尺度上的空间具有细小结构的线索。具体说来，贝肯斯坦与霍金发现，如果你在头脑中往黑洞视界上画一张跳棋盘，其中每个小格子的大小都是一个普朗克长度乘以一个普朗克长度（所以每个这样的“普朗克方块”的大小是$10^{-66}$平方厘米），那么黑洞的熵就等于填满整个黑洞表面所需要的小格子的数目。[4]到了这里我们就很难放过这样一个结论：每一个普朗克方块就是一个最小的空间基本单元，它所具有的熵就是最小单位的熵。这就意味着——即使只是在理论上——普朗克方块内什么都发生不了，因为任何一种活动都会带来无序度的提升，而那就会导致普朗克方块内的熵大于贝肯斯坦与霍金算出来的那一个单位的熵。于是又一次的，我们从一个全然不同的视角出发，最后得出了存在基本空间实体的认识。[5]

另外，对一位物理学家来说，某一空间区域中可能存在的熵的上限是一个临界的，几乎具有神圣意义的量。要明白为什么会这样，试着想象一下你正在跟一位行为精神病学家一起工作，你的任务是每时

每刻详细记录一群过度活跃的小朋友之间的相互影响。每天早上你都会祈祷上帝，希望今天小朋友能够表现得好点，因为小朋友们制造的麻烦越多，你的活就越麻烦。虽然直观上看理由很明显，但我们还是有必要说清楚：小朋友们的行为表现越混乱，你需要记录下来的内容就越多。宇宙就给了物理学家们一个这样的挑战。一个基本层面上的物理理论需要描述给定区域内一切正在发生的事——或者可能会发生的事，即使这事只有理论上的可能性。而且，就像记录小朋友们的活动那样，一个区域内所能容纳的混乱程度越高——即使只有理论上的可能性——物理理论所必须说明的内容就越多。因而，一个区域内所能容纳的最大熵就是一块简单但却锋利的试金石：物理学家们期望真正意义上的基本理论应该能够完美的匹配出任意空间区域的最大熵。这个真正的基本理论应该几乎不需要调节参数就能与大自然相符，其所能记录的最大混乱度应该正好等于一个区域所能有的最大无序度，既不多也不少。

问题在于，如果对特百惠容器的思考所得出的结论具有无限的有效性，那么一个基本理论就得有能力计算正比于任意区域体积的混乱度。但是这种思考却没将引力算在内，而一个基本理论又必须包括引力，于是我们知道了一个基本理论只要有能力计算正比于任意区域表面积的混乱度就足够了。从前面几段里给出的例子中，我们可以清楚地看到，对于比较大的空间区域，混乱度正比于面积会比正比于体积小得多。

因而，贝肯斯坦与霍金的结果告诉我们，在某种意义上，一个包括了引力的理论将比一个没有包括引力的理论简单些。包括了引力的理论所必须描述的“自由度”将更少——能够变化因而会对混乱度有贡献的东西更少。这一认识本身就非常有趣，而假如我们沿着这种思路更进一步的话，我们就将发现一些极为古怪的事情。如果任意给定的空间区域内的熵的最大值正比于空间的表面积而不是体积的话，那么或许真正基本的自由度——那些能够带来混乱度的特性——或许存在

于区域的表面而不是内部。或许，宇宙中真正的物理过程都发生在一张薄薄的环绕着我们的遥远表面上，我们所看到和体验到的一切只不过是这些过程的一个投影。或许，宇宙就像全息图一样。

这一想法非常古怪，但正如我们将要论及的那样，它最近得到了很多实质性的支持。

## 宇宙是一幅全息图吗

所谓的全息图实际上是一张二维的塑料板，但这张塑料板上所刻蚀的内容在适当的激光照射下会投影出三维的图像。[6]早在20世纪90年代，荷兰诺贝尔奖得主杰拉德·特霍夫特与弦论的另一个创造者莱昂纳德·萨斯金就曾提出，宇宙本身的运行模式可能就像全息图一样。这两位物理学家提出了一个惊人的想法：我们在日常生活的三维世界中所观测到的过去、现在和将来可能是发生在遥远的二维表面上的物理过程的全息投影。按他们新奇的观点，我们和我们所看所做的一切就像全息图一样。在柏拉图看来，普通的人类感知所感受到的只不过是真实的影子，全息原理承认这一点，只不过认为位置应该颠倒一下。影子——那些存在于低维表面的扁平家伙——才是真实的，而看起来结构更加丰富的高维事物（我们，以及我们周围的整个世界）反倒是影子那轻盈的投影。[96]

尽管这是一个相当不可思议的想法，并且我们也不知道这个想法究竟能在多大程度上帮助我们最终搞清楚时空，特霍夫特和萨斯金的所谓“全息原理”还是非常的有启发性。正如我们在上一节中所讨论过的，空间中某块区域所能容纳的最大熵由该区域的表面积而不是体积决定。于是自然就会想到，宇宙最基本的组成，其最基本的自由度——携带着整个宇宙的熵的那些东西，就像携有《战争与和平》的熵的书页——就应该在宇宙的边界面上而不是在宇宙的内部。我们在宇宙的“体内”——用物理学家常说的话即为“在体空间（bulk）中”——所体验的一切，实际上是由发生在边界面上的物理所决定的，正如我

们在全息投影中看到的一切是由刻蚀在塑料板信息编码决定的一样。物理定律如同宇宙的激光，照亮宇宙的真实过程——发生在远方薄薄的表面上的过程——并产生出我们日常生活的全息幻象。

我们还没能搞明白究竟怎样才能在真实世界中实现全息原理。困难之一在于我们对宇宙的传统描述，将宇宙想象成无限膨胀下去或者蜷曲回自身——就像球面或者电子游戏屏幕那样（参见第8章），不管怎样宇宙都不会有什么边界。这样一来，这个所谓的“边界全息面”究竟该被安置在哪就成了问题。而且，物理过程看起来不就是在我们的掌控之中吗？不就是发生在宇宙内部吗？怎么看也不像是连其位置都弄不清楚的边界的事物决定了体空间中发生的一切。难道全息原理要告诉我们的是，那种种一切尽在掌控的感觉和自主意识其实都是虚幻的？又或者我们应该将全息原理看成是某种对偶性？根据这种对偶性，每个人可以基于品味而不是物理选择熟悉的表述方式——在体空间中起作用的基本物理定律（这种体空间内的物理定律符合人类的直觉与感知），或者不熟悉的表述方式——在宇宙边界上起作用的基本物理定律，这两种视角彼此等价。关于这些重要的问题目前仍有很多争议。

1997年，在先前几位弦论学家的启迪下，阿根廷物理学家胡安·马达西纳取得了重大突破，出人意料地推动了对这些问题的思考。马达西纳的发现并没有直接回答全息原理在我们的真实世界中所扮演的角色问题。他只是按照传统的物理学家风格，找到了一个理想模型——一个假想的宇宙，在这个假想的宇宙中，关于全息原理的各种奇思妙想可以通过数学变得具体而精细。出于技术上的原因，马达西纳研究了一个具有4个空间维度和1个时间维度的假想宇宙，并且假定这个宇宙的所有维度都有相同的负曲率——你可以把这样的假想宇宙想象成图8.6（c）中那样的品客薯片。标准的数学分析表明这个五维时空有一个边界，[7]并且就像所有的边界一样，这个边界也比它所包裹的形状少1个维度：也就是说该边界有3个空间维度和1个时间维度

（和以前一样，我们还是很难想象高维形状的样子，所以你要是非得在脑海中想点什么东西的话，那就想象有一听番茄罐头——罐头盒中的三维番茄就可以类比于五维时空，而罐头盒表面的二维铁皮就类比于我们所要讨论的四维时空边界）。马达西纳又将弦论所要求的那些额外的蜷曲维度考虑进来，然后他满怀信心地提出：一个生活在这个假想宇宙中的观测者（番茄罐头中的观测者）所见证的物理可以由发生在这个宇宙边界上的物理（罐头盒表皮上的物理）全盘描述。

尽管这一研究工作所讨论的内容并不是真实世界的物理，但它却给出了将全息原理具体实现的第一个坚实且可数学化处理的例子。[8]依靠这种办法，人们对于将全息原理应用于整个宇宙多了一些概念上的认识。比如说，在马达西纳的工作中，物理定律的体空间描述和边界描述是完全等价的。两者之间不存在谁高谁低的问题。这里的精神实质就如同5种弦论之间的关系，体空间理论与边界理论彼此互为转译。这种特殊翻译的不平凡之处在于，体空间理论比等效的边界理论有更多的维度。另外，体空间理论中包括了引力（因为马达西纳是靠弦论得出这些结果的，而弦论是包括了引力的），而计算表明边界理论中则不包括引力。不过，在一个理论中提出的任何问题或者完成的任何计算都可以被翻译为另一个理论中的问题和计算。尽管不熟悉字典的人可能会认为两个理论中彼此对应的问题与计算一点关联性都没有（比如说，边界理论没有包括引力，因而在体空间理论中与引力有关的问题都得被翻译成边界理论中面目全非的无引力问题）；但是一个对于两种语言都很熟悉的人——同时为这两种理论的专家——却能看出两者之间的对应关系，并且会知道在不同理论中对应问题的答案以及对应计算的结果必定彼此符合。事实上，到目前为止完成的所有计算全部支持这一论断。

要想搞清楚这其中的种种细节显然是很困难的，但不要被这些细节遮住了要点。马达西纳的结果非常神奇。他在弦论的框架下，具体地实现了全息原理，虽然仅是一种与现实世界无关的理论上的讨论。

马达西纳证明了某种没考虑引力的量子理论其实是另一种包括了引力且多一个空间维度的量子理论的翻译，而且两者无法区分。而为了将这些思想应用于更加真实的宇宙——我们的宇宙——而展开的研究计划正在进行中，但由于技术上的复杂性，进展非常缓慢（马达西纳之所以选择了一个假想的研究对象就是因为他所选用的研究对象在数学分析上稍微容易一些，而更加接近于真实的例子则很难对付）。不管怎样，我们现在已经知道了弦论，至少在某种程度上，有能力支持全息原理的概念。而且，就像早前讲过的几何翻译的例子，这里讲到的全息原理也提供了一条时空不具有基本性的线索，除了时空的尺寸和形状可以在不同的理论之间来回翻译，空间维度的数目也可以在完全等价的不同理论之间改变。

越来越多的线索都指向同一个结论：时空的形式只是一个无关紧要的细节，在不同的物理理论体系下，时空的形式会发生改变，而不是真实性的一个基本元素。就像单词cat的字母数、音节以及元音全都不同于其西班牙语翻译gato一样，时空的形式——其形状、大小以及维数——也会在翻译过程中有所改变。对于任何一位运用某种理论来思考宇宙的观测者来说，时空看起来都是那么真实且不可或缺。而一旦这位观测者将其理论体系变变样子，用一个等价的翻译版再去分析，就会发现先前的真实与不可或缺已不再成立。因而，如果这些想法是正确的——我必须强调它们已被严格证明——那空间和时间的地位就会被强烈动摇。

在本章讨论过的所有例子中，我认为最有可能在未来的研究中扮演重要角色的将是全息原理。全息原理发端于黑洞的一个基本性质——黑洞的熵，而很多物理学家都会同意，对黑洞熵的认识需要坚实的理论基础。即使我们现有理论的细节需要有所改变，任何合理的引力理论也都必须为黑洞留有一席之地，于是从中得出的有关熵界的讨论将继续有效，全息原理也仍有用武之地。弦论自然而然地顺应全息原理——至少在可应用数学分析的例子中是这样的——是全息原理可

靠有效的另一个强有力的证据。我认为，不管未来有关空间和时间的基础的研究会把我们带到哪里，也不管等待我们的是弦论或M理论的哪种疯狂变种，全息原理都将继续保持其先导性概念的地位。

时空的组分

纵观全书，我们时不时地就会提到时空的超小尺度组分，尽管我们进行了很多有关其存在性的间接讨论，但到目前为止我们还没有谈过它们究竟可能是些什么。之所以如此实在不得已，因为我们根本不知道它们是些什么。或者，我可以这么说，一旦需要辨识时空的基本成分时，我们才发现我们根本不知道令我们信心满满的究竟是些什么。尽管这是我们思考过程中的一个主要障碍，但还是有必要从历史发展角度来仔细看看这个问题。

要是你能够对19世纪晚期的科学家做一个问卷调查，看看他们对物质的基本组成有些什么样的看法，你会发现根本找不到一个大家普遍认同的答案。一个多世纪前，原子假说还有很多争议；很多著名的科学家——欧内斯特·马赫就是其中之一——认为原子假说是错误的。而且，从原子假说被广为接受的20世纪前叶开始，科学家们就没有停止过更新原子论的图像，越来越多的基本组分被不断提出（比如说，先是提出了质子和中子，后来又提出了夸克）。沿着这条路一路走来，最新的一步就是弦论。但是因为弦论一直不能得到实验的证实（即使弦论被实验证实，那也并不意味着排除了其他有待发展的更高级理论），我们必须坦白承认对自然界中的物质的基本组成的研究尚未到头。

空间和时间被纳入现代科学理论可追溯到牛顿时代的17世纪，而对其微观组分的严肃思考则需要20世纪广义相对论和量子力学的发现。因而，从时间角度说，我们对时空的研究才刚刚起步；因而，不能给时空“原子”——时空最基本的组分——一个明确的说法并不是这个研究项目上的一大污点，甚至可以说是远远不是。我们所得到的已

经够多了——我们已经搞清了大量完全有别于日常经验的时空性质——这已经证明我们比一个世纪前进步了很多。对大自然的最基本元素——不论是物质的基本组分还是时空的基本组分——的研究，将会是我们在未来一段时间内所面临的强大挑战。

对于时空来说，在寻找其基本组分方面目前存在两个大有希望的方向。其中之一为弦论，而另一个则是所谓的圈量子引力理论。

弦论的方案要么令你直觉上觉得很不错，要么令你感到非常困惑，全看你究竟把它想得多难。当我们说到时空的“结构”时，我们的话中已经暗含了弦论的提案，或许时空就是由弦编织起来的，就像衬衫是由线编织起来的那样。这就好比将大量的线按照一定的模式连接起来就产生了衬衫的结构，或许将大量的弦按照一定的模式连接起来就产生了我们通常称之为时空结构的东西。物质，比方说你和我，则可归结为振动的弦的再次聚合——就像明快的音乐是由一个个单调的声音组成或是漂亮的绣花是由朴素的材料织成，而物质还将在时空的弦织成的框架中移动。

我认为这是一个引人入胜的想法，可惜还没人能够将它转变成精确的数学语言。我只能告诉你，它实在是比你想象的还要难对付。比方说，要是你的衬衫被完全扯烂了，就会剩下一堆线头——虽然视环境不同，你可能会感到难堪或者恼火，但不会感到有什么神秘之处。而讨论的对象变成弦的话——在我们讨论的这个想法中，弦就是用来编织时空的线——那就颇费脑力了（至少对我是这样）。我们究竟该怎样看待这“堆”从时空布片上扯下来的弦？或许更为切中要害的说法应是，我们究竟该怎样看待还没有缝合成时空之布的弦呢？我们可能会简单地将这些弦类比成织成衬衫的线——就将弦想象成需要编织起来的原材料——但要这么想的话就丢掉了一个极其重要的细节。在我们勾画的图像中，弦是在时空中振动。但是，若时空之布是由弦有序编织起来的话，就根本没有空间，也没有时间。在我们要讨论的这个想法中，空间和时间的概念根本就没有意义，除非数不胜数的弦交织

起来编成时空。

因而，要想令这个方案有意义，我们还需要一个理论框架来描述弦。在这个框架下，我们用不着从一开始就假定弦在默认存在的时空中振动。我们需要一个无空无时的弦论体系，在这个体系中，时空来自于弦的集群效应。

尽管我们在这个方向上已经取得了一些进展，但还没人能够给出这样一套无空无时的弦论体系——物理学家将具有这种属性的体系称为不依赖于背景的体系（使用这样术语的原因在于，物理学家们将时空视作一个不严格的背景，以区别于发生于其中的物理事件）。与之相反的是，本质上所有的方法都认为，弦是在时空中移动以及振动，而时空的概念则是“用手”放到理论中的；时空并非脱胎于理论本身——在物理学家设想的不依赖于背景的体系中，时空就应该来自于理论本身——而是由理论学家添加到理论中去的。很多研究者将不依赖于背景的体系的发展看做是弦论所面临的未解决问题中最重大的一个。不依赖于背景的体系的发展不仅会对我们关于时空起源的认识有所启发，还可能是我们在第12章最后遇到的重大问题——理论目前还不能挑选出额外维度的几何形式——的解决工具。一旦其有关任意给定时空的数学公式问题得以解决，我们的研究就可以进行下去，弦论也将有能力勘察所有的可能性，并从中找出正确的那个。

“把弦当做织起时空之线”这个方案所面临的另一大难题在于，如我们在第12章中讲过的，弦论中除了弦还有其他东西。这些其他组分在时空的基本组成中扮演的又是什么角色？这一问题在膜世界方案中变得尤为尖锐。如果我们所感受到的三维空间是一个3-膜，那么这个膜究竟是本身即不可再分呢？还是由理论中的其他组分构成呢？比方说，膜究竟是由弦构成的呢？还是说膜和弦都具有基本性呢？或者我们还需要考虑其他的可能性吗？比方说膜和弦其实都是由更基本的元素构成这种可能性。这些问题就是现今的研究前沿。但既然这最后一章所关注的只是迹象与线索，就让我讲一个引起诸多关注的相关思

想。

稍早前，我们讲过人们可以从弦论或M理论中找到的各种膜：1膜、2-膜、3-膜、4膜，如此等等。尽管之前我没特意强调过，但理论中也可以有0膜——没有空间延展性的组分，就像点粒子那样。0膜看起来与整个弦论或M理论格格不入，弦论或M理论本来就是要从点粒子理论的框架下解脱出来，以调和量子引力中剧烈的波动。然而，0膜就跟它那如图13.2所示的高维兄弟们一样，附着于弦而来，因而其相互作用完全受弦掌控。所以，毫不奇怪的是，0膜不会有同传统的点粒子一样的行为；而且，最重要的是，0膜彻底的参与到超微观时空涨落的延展及减弱中；0膜并不会重新引入困扰着同样试图融合量子力学和广义相对论的点粒子方案的致命缺陷。

事实上，罗格斯大学的汤姆·班克斯，得克萨斯大学奥斯丁分校的威利·费施勒，以及斯坦福大学的莱昂纳德·萨斯金和史蒂芬·森克尔提出了一个弦论或M理论，在这个理论中，0膜才是基本组分，它们可以组合到一起生成弦及其他高维的膜。这一提案，所谓的矩阵理论（MatriXTheory）——M理论中的M的又一层意义——引发了一股后续研究热潮，但是其中所涉及的数学如此之难，以至于目前为止科学家们还没法将这个理论完善起来。不过，物理学家们在这个理论框架下努力完成的计算看起来是支持这一提案的。如果矩阵理论正确，那就意味着一切事物——弦、膜，甚至时空本身——都是由0膜的恰当集合构成。这一理论的研究前景令人振奋，研究者们持谨慎的乐观态度，认为未来几年的进展将使人们了解该理论的有效性。

到目前为止，我们一直沿着弦论研究时空结构的道路前进。但正如我提到过的，还有另一条道路，它来自于弦论主要竞争对手圈量子引力。圈量子引力出现于20世纪80年代，是另一种有希望将广义相对论和量子力学融合起来的方案。我不打算详细讲解该理论了（如果你有兴趣，可以看看李·斯莫林的优秀作品——《量子引力的3条道路》），但会提一下对我们目前的讨论特别有意义的几个关键点。

弦论和圈量子引力全都宣称自己实现了人们长久以来寻找引力的量子理论这个目标，但两者达到这一目标的方式却完全不同。取得成功的粒子物理学几十年来的传统就是寻找物质的基本组成，弦论发轫于粒子物理；对于弦论的最早研究者来说，引力最多只是一个稍远的次要关注点。与之相反的是，圈量子引力则脱胎于一个与广义相对论紧密相连的传统；对于这一方法的大量研究者来说，引力一直都是主要焦点之所在。一言以蔽之，弦论学家从小（量子理论）开始，进而包围大（引力），而圈量子引力的追随者则从大（引力）开始，进而包围小（量子理论）。[9]事实上，正如我们在第12章中看到的那样，弦论最初是作为在原子核内起作用的强核力的量子理论发展起来的；后来人们才偶然认识到，弦论实际上可以将引力纳入其中。而另一方面，圈量子引力从爱因斯坦的广义相对论出发，试图将量子力学纳入其中。

弦论与圈量子引力这两个理论彼此首尾对应。在某种意义上，一个理论所取得的重大成就将是另一个理论的重大挫折。比方说，弦论试图用振动着的弦的语言来融合所有的力与所有的物质，其中也包括引力（避开圈方法的完整统一）。传递引力的粒子引力子，只不过是弦的一种特殊振动模式，因而弦论可以很自然的用量子力学的语言描述这些基本的引力丛如何运动如何相互作用。然而，正如我们刚刚提到过的，弦论体系目前主要的缺陷即在于预先假定了弦运动以及振动于其中的时空的存在。相反，圈量子引力的主要成就——这一点令人印象深刻——则在于其并不假定时空的存在。圈量子引力就是一个不依赖于背景的体系。然而，从这个极其陌生的以无空、无时为起点的理论中，抽取出普通的空间和时间，以及在大尺度上获得人们熟悉且已取得成功的广义相对论的性质（用目前的弦论体系可能很容易做到），却远非易事。而且，与弦论相比，圈量子引力在对引力子动力学的理解上几乎没有前进半步。

也有一种调谐两者的可能性，即弦论的追随者与圈量子引力的拥

逭所构建的可能是同一种理论，只不过出发点相差巨大。这两种理论都与圈有关——在弦论中，圈就是弦圈；而在圈量子引力中，不用数学语言实际上很难描述圈指的到底是什么，但大体说来指的是空间的基本圈——可能意味着两者之间有某种联系。而对于少数在两个理论中都可以很好的研究的问题，比如说黑洞的熵，两个理论给出的结果完全相符，这个事实也进一步支持了两者可能是同一理论这种可能性。[10]而且，在时空组分的问题上，两者都提出时空可能有某种微小的结构。我们已经知道了弦论所提供的有关这一问题的线索；而来自于圈量子引力的线索虽然复杂但是更具说服力。圈方面的专家已经证明，圈量子引力中的无数圈可以彼此交织在一起——就像细细的羊毛卷可以编织成毛衣——生成在大尺度上看起来如同时空中的区域一般的结构。最令人欢欣鼓舞的是，圈方面的专家计算出了这种空间表面可以容许的面积。什么意思呢？比方说你可以找到1个电子、2个电子或者202个电子，但你就是不能找到1.6个电子或其他分数个电子。圈量子引力的专家们用计算证明，他们找到的空间的表面积只能是1普朗克长度的平方或者2普朗克长度的平方，又或者202普朗克长度的平方，总之不会是分数个普朗克长度的平方。再次重申，这是一个空间——就像电子——具有分立，具有不可再分的结构的重要理论线索。[11]

若要我斗胆猜一猜未来的发展的话，我估计圈量子引力学家们开发的不依赖于背景的体系会为弦论学家借用，为不依赖于背景的弦论体系开道铺路。我猜，这星星之火可能会燃起第3次弦论革命；乐观估计的话，这第3次弦论革命可能会解开很多深奥的谜题。这样的进展可能会为有关时空的漫长故事画上圆满的句号。在较早的章节中，我们有关空间、时间以及时空的观点就像钟摆一样，在相对论者与绝对论者之间摆来荡去。我们曾问道：空间是具体的还是抽象的？时空是具体的还是抽象的？而人类在过去几个世纪的思考，带给我们的却是各种彼此不同的观点。我相信，一个能够在实验上验证的，可以将广义相对论和量子力学联合起来的不依赖于背景的理论将会为这些问

题提供令人满意的答案。由于这样的理论所具有的不依赖于背景的特性，这个理论中的不同部分会彼此关联，但这个理论中不会有从外面放进去的时空，理论中的各个元素将找不到一个搭好的背景舞台。重要的将只是相对关系——这是一个具有莱布尼茨或者马赫这样的相对论者所持有的精神实质的答案。那么，由于该理论中的组分——弦、膜、圈，或者在未来的研究中发现的其他什么东西——结合起来产生了我们熟悉的大尺度时空（不论是我们的真实时空还是在假想实验有用的理论例子），因而时空重新变得具体起来。这一点类似于我们之前对广义相对论的讨论：即使在一个空空如也、平直、无限大的时空中（一个有用的假想例子），旋转着的牛顿的桶中的水面也会凹陷。关键之处在于，时空和更加切实的物质之间的界限会变得模糊，因为两者都来自于理论——在基本层面上无空无时的理论——中更加基本的元素的适当集合。如果最终的答案真是这样，那莱布尼茨、牛顿、马赫以及爱因斯坦就都要宣布分享胜利了。

太空的里外

预测科学的未来既有趣又富于建设性。预测科学的未来就是要在更为宽广的背景中看待我们今日之事业，特别是我们为之努力的目标。但要说到猜想有关时空自身的研究的前景，那就难免会带有某种神秘色彩了：因为我们所要思考的乃是主宰着我们对真实性的理解的概念。再次重申，不论未来有何发现，空间和时间无疑都将继续扮演我们个人体验的载体这个角色；空间和时间，无论时光如何流逝，都将一成不变。将会持续发生改变的，将会发生不可思议的变化的，乃是我们对空间和时间所提供的框架——从实验的真实性角度说，空间和时间为我们搭建的舞台——的理解。几个世纪的思索之后，空间和时间仍然是我们最熟悉的陌生人。空间和时间从不介意现身于我们的日常生活，但却始终巧妙地将自己的基本组成掩藏于其无处不在的影响力之下。

在过去的20世纪，通过爱因斯坦的两个相对论，通过量子力学，

我们已经对先前那些隐藏起来的空间和时间之特性不再陌生了。时间的变慢，同时性的相对性，其他的时空片，作为空间和时间蜷曲和弯曲原因的引力，真实性的概率性，长程量子纠缠，所有的这些问题早已超出了19世纪时世界上最棒的物理学家的研究计划。但实验数据与理论诠释可以证明，这些问题的确存在。

在我们这个时代，我们也有能使自己震撼的奇思妙想：暗物质和暗能量看来必是宇宙的主要成分。爱因斯坦的广义相对论所预言的引力波——时空结构中的波纹——在未来的某一天可能会使我们有机会洞察时间。遍布整个空间的希格斯海，一旦得到实验的证实，将会帮助我们搞清楚粒子如何获得质量。暴涨式膨胀，有可能解释宇宙形状，解开宇宙在大尺度上的各向同性之谜，并为时间之箭点明方向。弦论用圈和能量片代替点粒子，并且承诺实现爱因斯坦之梦——用一个单独的理论统一所有的粒子和所有的力。弦论的数学中自然出现的额外空间维度，有可能在未来10年内通过加速器实验得以验证。在膜世界理论中，我们的三维世界被假定为漂浮在高维时空中的众多世界中的一个。甚至就连时空本身也成了问题，空间和时间的微观结构本身就是由更基本的无空无时的实体构成。

接下来的10年间，前所未有的强大加速器将为物理学家们提供急需的大量数据；很多物理学家都相信，从计划中的高能实验中收集的数据将验证很多我们讲过的新理论。我也无法免于这种狂热，也在急切地期盼着结果。在我们的理论能够与可观测量，与可检验的现象联系起来之前，它们一直徘徊在危险的边缘——它们是否能同真实世界关联起来还没人知道。新的加速器能将实验与理论更加紧密地联系起来；我们这些物理学家希望新的加速器能将大量的想法变成扎实的科学理论。

但还有另外一种令我觉得奇妙到无与伦比的想法，虽然其胜出的机会非常之小。在第11章中，我们讨论过如何在清晰的夜空中看到小小的量子涨落效应，其根据就是宇宙的膨胀会将这小小的量子效应充

分拉伸，使之聚团，成为恒星和星系的种子（还记得我们做的那个类比吗？气球上的涂鸦随着气球的膨胀而在气球表面延展开来）。这个例子展示的就是通过天文学观测探索量子物理的办法。而且我们还可以进一步深入下去。或许宇宙的膨胀会使更小尺度上的过程的印记或特征——有关弦之物理，或者更普遍意义上的引力，或者超微观尺度上时空自身的微小结构——得以放大，并通过复杂但却可以通过天文学办法观测的方式使其影响得以彰显。或许，宇宙已经将其微小纤维或结构在天空中清楚地画了出来，我们需要做的只是对其特征加以识别。

对深层次物理前沿理论的检验可能需要能够重新创造出大爆炸之后从未再现过的猛烈瞬间的强力加速器。但对我来说，比起确证我们有关超小尺度上的理论——我们有关超小尺度上的空间、时间，以及物质组成的理论——没有什么其他的理论能更具诗情画意，也没有什么其他结果更为优雅，更不会有什么其他的统一理论会更加完备了。

# 注释

第1章

1.物理学家阿尔伯特·迈克尔逊于1894年在致芝加哥大学的Ryerson实验室的献辞中引用过开尔文勋爵（见D.Kleppner, Physics Today，1998年11月）。

2.罗德·开尔文，“Nineteenth Century Clouds over the Dynamical Theory of Heat and Light”，Phil.Mag.Ii 6th series，1（1901）。

3.阿尔伯特·爱因斯坦、内森·罗森和鲍里斯·波多斯基，Phys.Reu.47，777（1935）。

4.阿瑟·爱丁顿爵士，The Nature of the Physical World（Cambridge, Eng.：Cam-bridge University Press，1928）。

5.就像第6章注释2中更加详尽解释的那样，这样的说法有点过头，因为有些与相对奇怪的粒子（比如K介子和B介子）有关的例子表明，所谓的弱核力并没有平等的对待过去和未来。然而，从我以及其他思考过该问题的人的角度看来，由于这些粒子并没有在决定日常物体性质中起关键作用，所以它们不可能对解释时间之箭的谜团有什么重要作用（虽然，我不得不补充一句，没有人能肯定）。因此，虽然从技术水平上来说，这样描述有点过头，但我可以肯定，假设这些定律平等的对待过去和未来并不会犯什么大错——至少在解释时间之箭的谜团时是这样。

6.Timothy Ferris, Coming of Age in the Milky Way（New York：Anchor，1989）.

第2章

1.伊萨克·牛顿，《自然哲学之数学原理》。Sir Isaac Newton's Mathematical principle of Natural Philosophy and Hissystem of the world，译者，A.Motte与Florian Cajori（Berke-ley：University of California Press，1934），vol.1，p.10。

2.同上，P.6。

3.同上。

4.同上，P.12。

5.阿尔伯特·爱因斯坦为Max Jammer的著作Concepts of SPace：The Histories of The-ories of Spacein Physics（New York：Dover，1993）写的序言。

6.A.Rupert Hall, Isaac Newton, Aduenturer in Thought（Cambridge, Eng.：Cam-bridge University Press，1922），p.27.

7.同上。

8.H.G.Alexander编辑，The Leibniz-Clarke Correspondence（Manchester：Manchester University Press，1956）。

9.在那些反对空间可以独立于居于其间的物质而存在的人中，我以莱布尼茨为代表，但还有许多其他人也尽力维护相同的观点，其中

克里斯蒂安·惠更斯和伯克莱主教较为知名。

10.参见，比如说，Max Jammer, P.116。

11.V.L.列宁，Materialism and Empiriocriticism：Critical Comments on a Reactionary Philosophy（New York：International Publications，1909）。为Materializm'i Empiriokrititsizm'：Kriticheskia Zametki ob'Odnoi Reaktsionnoi Filosofii（Moscow：Zveno Press，1909）的英文第2版。

第3章

1.对于受过数学训练的读者而言，这4个方程是：

∇·E＝ρ/ε0，∇·B＝0，∇×E＋αB/αt＝0，∇XB-ε0μ0αE/αt＝μ0J，

其中E，B，ρ，J，ε0，μ0分别表示电场强度、磁场强度、电荷密度、电流密度、自由空间的介电常数和自由空间的磁导率。如上所示，麦克斯韦的方程将电磁场的变化率与电荷、电流联系起来。不难看出这些方程暗示着电磁波的波速为1/ε0μ＼0，而这实际上就是光速。

2.关于这些实验在爱因斯坦发展狭义相对论的过程中所起的作用还存在争议。在爱因斯坦的传记Subtle Is the Lord：The Science and the Life of Albert Einstein（Oxford：Oxford University Press，1982）pp.115-119中，亚伯拉罕·派萨认为，从爱因斯坦后几十年的陈述中可以看出他承认迈克尔逊莫雷实验的结果。Albrecht Fölsing在Albert Einstein：A Biography（New York：Viking，1997）pp.217-220中也写道，爱因斯坦承认迈克尔逊莫雷实验的结果以及早期寻找以太证据时所做实验的无效结果，比如阿曼德·菲佐的工作。但Fölsing和许多科学史家们都认为这些实验在爱因斯坦思想的形成中起

了次要的作用。爱因斯坦首先是受数学的对称性、简易性和神秘的物理直觉引导的。

3.如果我们想要看到任何事物，光就不得不到达我们的眼睛；同理，如果我们要看到光，光本身也要先到达我们的眼睛。因此，当我说巴特看到飞驰的光时，这只是一种简略的说法。我们可以想象巴特有一队助手，都以和巴特相同的速度运动着，并且分布于他和光束所走路径的不同位置。他们不断给巴特更新信息：光走了多远以及光到达他们的时间。于是，基于这些信息，巴特就可以计算光比他快多少。

4.爱因斯坦对于时间和空间的见解源于狭义相对论，有许多基本的数学推导。如果你感兴趣的话，可以看一下《宇宙的琴弦》的第2章（在该章的注释中有很多数学细节）。埃德温·泰勒和约翰·阿奇博尔德写有一本更加专业但却非常清楚的报告Spacetime Physics：Introduction to Special Relatiuity（New York, W.H.Freeman & Co.，1992）。

5.以光速运行时，时间将停止是一个非常有趣的概念，但不要对这做过多解释。狭义相对论表明没有物体可以达到光速：一个物体运动得越快，我们就越难使它的速度增大。由于小于光速，我们将不得不给物体一个无限大的推力使其运动得更快，但这是我们所不能做到的。这样看来，“永恒的”（不受时间影响的）光子的说法仅限于无质量的物体（光子就是一个例子），因此“永恒之物”不过是一些可以达到标准的粒子而已。如果我们想要知道狭义相对论是如何影响我们的时间感受的，不妨想象一下当我们以光速运动时宇宙发生了什么变化，这是一个非常有趣且有益的游戏，最终我们的焦点将集中在物体比如我们将会发生哪些变化上。

6.见亚伯拉罕·派萨的Subtle Is the Lord pp.113-114。

7.为了使描述更加精确，我们将水面处于凹形的情形定义为水在旋转，反之则水没在旋转。从马赫式的角度来看，在一个空的宇宙里没有旋转的概念，因此水面总是平的（或者，为了避免出现没有重力作用于水的问题，我们说连接两块石头的绳子的张力总是松弛的）。这里的陈述即为，通过对比，狭义相对论中有旋转的概念，即便在真空的宇宙中也是如此，因此水面可以是凹的（连接两块石头的绳子的张力是紧绷的）。从这种意义上来说，狭义相对论背离了马赫的观点。

8.Albrecht Fölsing, Albert Einstein（New York：Viking Press，1977），pp.208-210.

9.数学功底不错的读者将会发现，如果我们通过选择单位，使光速取每单位时间走单位距离的形式（如每年1光年，每秒1光秒，其中1光年约6万亿英里，1光秒约186000英里），那么光将以45度的光线穿越时空（因为这样的对角线满足单位时间内走单位空间，两个时间单位内走两个空间单位，等等）。因为没有物体能够超越光速，所以任何物体在某一时间段走过空间不可能比光更多，因此该物体穿越时空的路径一定与图的中心线（该线起于面包皮，穿越了面包的中心，最后止于面包皮）呈一定角度，并且该角度要小于45度。而且，爱因斯坦指出，以速度u运动的观测者的时间片——该观测者处于某一时刻时的全部空间——都符合一个方程为简易起见，假设只有一个空间维度$t_{运动}=\gamma[t_{静止}-(\upsilon/c^2)x_{静止}]$，其中，$\gamma=(1-\upsilon^2/c^2)^{-1/2}$，c表示光速。当单位c＝1时，u＜1，因此对于运动的观测者的时间片而言——$t_{运动}$取定值的位置——（$t_{静止}-\upsilon x_{静止}$）＝定值。这样的时间片与静止的时间片（$t_{静止}$＝定值的位置）之间存在一定角度，因为u＜1，所以它们之间的角度小于45度。

10.对于数学功底不错的读者而言，闵科夫斯基时空的最短路径——两点之间最短的时空长度——是不依赖于任何坐标或参考系的几何实体。它们是内在的、绝对的、几何的时空性质。明确地说，用闵

科夫斯基的度规标准来看，（类时）最短路径是直线（该线与时间轴的角度小于45度，因为不可能出现比光速更快的速度）。

11.重要的是所有的观测者，不管他们的运动状态如何，都将在这一点上达成一致。我们的描述中已间接说明了这一点，但它更值得我们直接说明。如果一件事情是另一件事情的原因（我扔了一块石头结果砸碎了窗户的玻璃），则所有人都会同意原因发生在结果之前（所有人都认为我先扔石头，然后窗户的玻璃碎了）。对于数学功底不错的读者而言，用数学的示意图描述将不难看出这一点。如果事件A是事件B的原因，那么从A到B的线条与每个时间片（事件A发生时观测者所看到的时间片）相交的角度将大于45度（空间轴与AB线之间的角度——轴位于给定的时间片——大于45度）。举个例子来说，如果A和B发生在空间的同一位置（橡皮筋绑着我的手指[A]结果我的手指变白[B]），那么连接AB的线条将与时间片呈90度的角度。如果A、B发生在空间的不同位置，从A到B发生的任何事情所造成的影响（我扔的石头从射击点到达窗户）都将比光速慢，这就意味着角度将不同于90度（无速度时的角度），也不会小于45度——也就是说，与时间片（空间轴）的角度将大于45度（本章的注释9中曾说明因光速的限制，这类运动轨迹最多呈45度）。现在我们再来看看注释9，处于运动状态的观测者的时间片与静止的观测者的时间片呈一定角度，但这个角度总是小于45度（因为两个观测者之间的运动的速度总是小于光速）。因为因果相关事件的角度总大于45度，观测者（他的速度肯定小于光速）的时间片，不可能先遇到结果再碰到原因。对所有的观测者而言，先因后果。

12.如果电磁感应比光速还快，上一条注释中所说的先因后果的观点将受到挑战。

13.伊萨克·牛顿，Sir Isaac Newton's Mathematical principles of Natural Philosophy and His System of the world，翻译A.Motte与Florian Cajori（Berkeley：University of Cali-fornia Press，

1962），vol.1，p.634。

14.因为在地球表面不同位置所受到的引力不同，所以一个双臂伸展开的、处于自由落体运动的观测者也可以探测到剩余的引力的影响。也就是说，如果观测者下落时释放了两个棒球——一个从伸展开的右胳膊扔出，另一个从左胳膊扔出——每个都朝地心的位置下落。这样，从观测者的角度来看，他将竖直下落到地心，而他右手释放的球将竖直向下运动并稍微向左边一点，左手释放的球将竖直下落并稍微向右边一点。经过仔细的测量，观测者将发现这两个球之间的距离慢慢减小了；它们彼此向对方移动了一点儿。之所以是这样关键就在于，棒球从空间中两个略为不同的位置释放，所以它们自由下落到地心的路径也略有不同。因此，对爱因斯坦观点更准确的描述应为：一个物体的空间体积越小，它通过自由下落消除引力就越彻底。虽然这是理论中非常重要的一点，但在我们的讨论中完全可以放心的将其忽略。

15.要想更详细、更通俗地了解一下广义相对论对空间和时间的蜷曲的解释，可以看看《宇宙的琴弦》的第3章。

16.对于有过数学训练的读者而言，爱因斯坦的方程为 $G_{\mu\upsilon}=(8\pi G/c^4)T_{\mu\upsilon}$，其中等式左边是用爱因斯坦张量表示的时空曲率，右边是用能量动量张量表示的宇宙中物质和能量的分布。

17.查尔斯·迈斯纳、基普·索恩和约翰·阿奇博尔德·惠勒合著的Grauitation（San Francisco：W.H.Freeman and Co.，1973），pp.544-545。

18.在1954年，爱因斯坦给同事的信中写道：“实际上，人们应当不要再谈马赫原理了（在亚伯拉罕·派萨的《上帝是微妙的》中被引用，p.288）。”

19.就像前文中提到的，连续几代人把该想法归功于马赫，尽管他在自己的著作中并没有以这种方式明确地表达出来。

20.这儿需满足的一个条件是：那些在宇宙起源时距离我们很远，以至于它们的光——或引力作用——没有足够的时间到达我们的物体，对我们感受到的引力不会产生影响。

21.专业的读者可能会意识到这种说法，从学术的语言来看，有点说过头了，因为对于广义相对论而言存在非平凡的（也就是说，非闵科夫斯基空间）真空。我在这里利用了一个简单的事实做了简化，即没有引力存在的情况下，狭义相对论可被看做广义相对论的一种特殊情况。

22.为了平衡，我们来看一下不同意这一结论的物理学家和哲学家的说法。即使爱因斯坦放弃了马赫原理，在最近30年间马赫原理仍自有其存在价值。现在关于马赫原理有各种各样版本的诠释，比如，一些物理学家认为广义相对论实际上从根本上包含了马赫观点；只是时空并不具有它本可以具有的某些特殊形状——例如像真空宇宙的无限大平直时空。或许，他们提出，其他一些并不现实的时空——其中满是恒星和星系之类——实际上满足马赫原理。其他人则重塑了马赫原理的体系。在他们的体系中，问题不再是物体（比如说被绳子系着的两块石头或装满水的桶）在真空宇宙是怎样的，而是各种不同的时间片——各种不同的三维空间几何——究竟是怎样通过时间与其他时空片联系起来的。对有关这些想法的现代思考极具参考价值的是Mach's Prin-ciple：From Newton's Bucket to Quantum Grauity（《马赫原理：从牛顿的桶到量子引力》）Julian Barbour与Herbert Pfister编辑（Berlin：Birkh?user，1995），该书收集了与此问题有关的一些论文。有趣的是该书包含了一项民意测验：40名物理学家和哲学家对马赫原理的看法。大多数人（90%多）认为广义相对论并不完全符合马赫原理。另一本以明显的超马赫式观点对此问题进行讨论，并在一定程度上适合于普通读者阅读的，这本优秀且极其有

趣的书是Julian Barbour的The End of Time：The Next Reuolution in Phys-ics（Oxford：Oxford University Press，1999）。

23.喜爱数学的读者在知道爱因斯坦认为空间不能独立于它的度规（描述时空距离关系的数学工具）而存在的观点后会感到很受启发，如果有人移走了每一样东西——包括度规——时空将不再是某种实体。当提到“时空”的时候，我指的都是一个拥有可解爱因斯坦方程的度规的流形，因此我们所得出的结论，用数学语言说就是，度规时空才是某种实体。

24.Max Jammer, Concepts of space, p.Xvii.

第4章

1.更确切地说，这看起来像是个可以追溯到亚里士多德的古老概念。

2.就像本书后面将要讨论的，在许多领域中（像大爆炸和黑洞）还存在着大量未解之谜，这些谜题至少可以部分的归结为：体积过小密度过高，从而造成爱因斯坦的优雅理论无法适用。因此，这里的描述适用于大部分情况，但已知定律无法应用的极端情况除外。

3.本文的最早读者中有一位擅长伏都魔咒，他告诉我，有些东西可以根据施魔咒者的意念——也就是灵魂——从一个地方被传送到另一个地方。因此，我这个充满幻想的非定域性例子——决定于你是否会伏都魔咒——可能是错的。不管怎么说，你搞清楚其中的思想就行了。

4.为了避免混乱，我再强调一下我刚才说的，“宇宙并非定域性的”，或者“我们在此地做的事情会与彼地的事情发生联系”，我并不是指能用瞬间的意念控制远在千里之外的事物。换句话说，我指的是相隔很远的地点（这两个地点距离非常远以至于都没有足够的时间让光

从一个地点到达另一个地点）发生的事情之间的联系——通常以测量结果之间的联系的形式表示。因此，我指的是物理学家们所说的非定域关联。乍看之下，这样的关联可能不会使你特别惊奇。如果有人送给你一只盒子，里面有一只手套，而配套的另一只手套送给了你远在千里之外的朋友，你们其中任何一人打开各自的盒子看到的手套将与另一人的存在关联：如果你看到的是左手戴的那只，你朋友将看到右手那只；如果你看到右手那只，你朋友将看到左手那只。显然，这种关联并没有什么神秘之处。但是，就像我们逐渐要讨论的，量子世界中这种明显的关联就是一种与众不同的性质了。这就好像你有一副“量子手套”，一只左手的，一只右手的，当观察或者相互作用时，它们就会表现出明显的偏手性。诡异之处在于，虽然观察时每只手套都会随机选择偏手性，但即使它们被相隔很远，这两只手套也会一前一后地选择相应的偏手性：如果一只选择左边，另一只就会选择右边，反之亦然。

5.量子力学预言的微观世界与实验观测完全符合。就凭这一点，它就得到了大家的普遍认同。然而，本章中将要讨论的量子力学的具体特点，完全不同于常识，而且由于理论的数学公式不同（关于理论如何填补微观现象和宏观测量结果之间的空白的不同公式），人们就如何诠释理论（然而，理论可以从数学上解释各种各样的令人迷惑的数据）的各种特点并没有达成共识，包括非定域问题。在本章中，我持一种特殊的观点，我觉得它是建立在现代流行理论的理解和实验结果的基础上的最令人信服的观点。但在这儿我得强调一下并不是所有的人都同意这个观点，在后面的注释中详尽的解释完该观点之后，我将简要地介绍一下别人的观点并介绍更多相关的参考资料。我也得强调一下，在稍后的讨论中，实验将与爱因斯坦的信念（他认为实验数据只能用粒子具有明确但隐藏起来的性质来解释，而该性质没有任何非定域纠缠）相矛盾。不管怎样，这个观点的失败只能排除一种定域宇宙。它不能排除粒子有明确但隐藏起来的性质的可能性。

6.对于数学功底不错的读者而言，我们来看看这种描述可能导致的误解。对于多粒子系统而言，概率波（用标准术语的话，就是波函数）与刚才所讲的内容有本质上一样的诠释，但概率波被定义为粒子在位形空间的函数（对一个单独的粒子而言，位形空间与真实空间同构，但如果是N粒子系统，则它就变成3N维的了）。这对于我们思考波函数是一个真实的物理实体还是仅仅是一个数学工具非常重要，因为如果站在前者的立场，我们就需要认同位形空间的实在性——第2章、第3章主题的有趣变异。在相对论量子场论中，场可以在通常的四维时空中定义，但也有一些用得比较少的体系，使用了推广的波函数——定义在更为抽象的空间“场空间”上的所谓波泛函。

7.我在这儿提到的实验是光电效应实验，在该实验中，光照在各种金属上使金属表面发出电子。实验学家们发现，光的强度越高，发射出去的电子就越多。而且，实验还表明每个发射出去的电子的能量是由光的颜色——频率——决定的。爱因斯坦认为，如果光是由粒子组成的话，这很容易理解，因为光强越大就意味着光束中光的粒子（光子）越多——而且光子越多，撞击金属表面的机会就越多，从而使得金属发出的电子也越多。进一步来看，光的频率会决定每个光子能量，因此每个发射出去的光子的能量将精确的与数据相符。光子的粒子性最终在1923年被亚瑟·康普顿通过电子和光子的弹性散射实验证实。

8.索尔维国际物理讨论会，《第5届索尔维国际物理讨论会会议论文集》（Rapport et discussions du 5ème Conseil）（Paris，1928），pp.253ff。

9.Irene Born, trans.，《博恩和爱因斯坦书信集》（New York：Walker，1971），p.223。

10.亨利·斯坦普，Nuouo Cimento 40B（1977），191-204。

11.戴维·玻姆是20世纪量子力学领域最富有创造性的科学家之一。他于1917年出生于宾夕法尼亚州，曾是伯克利罗伯特·奥本海默的学生。在普林斯顿大学教书时，他曾被众议院非美活动调查委员会传唤，但他拒绝在听证会上作证。后来他离开美国，先后在巴西的圣保罗大学（S?o　Paulo），以色列的理工大学（Technion），伦敦大学的伯克贝克（Birkbeck）学院任教授。1992年他在伦敦逝世。

12.当然，如果你等的时间足够长，你对一个粒子做的事情，从原理上讲，会影响另一个粒子：一个粒子可以发出某种信号警告另一个粒子它正被测量，而且该信号也会影响接收粒子。然而，由于没有信号会比光的速度更快，所以这种影响不会即刻发生。现在讨论的关键点在于，我们测量粒子绕某一给定的轴自旋时，会发现其他粒子也绕该轴自旋。因此，粒子之间的任何一种“标准”通信——光速或亚光速通信——都不能说明问题。

13.在本节和下一节中，我对贝尔发现的归纳，受到David Mermin的美妙文章的启发：“Quantum Mysteries for Anyone，”Journal of Philosophy 78，（1981），pp.397-408；“Can You Help Your Team Tonight by Watching on TV？”收录在Philosophical Consequences of Quantum Theory：Reflections on Bell's Theorem, James T.Cushing与Ernan McMullin编辑。（University of Notre Dame Press，1989）；“Spooky Action at a Distance：Mysteries of the Quantum Theory”发表于The Great Ideas Today（Encyclopaedia Britannica, Inc.，1988），全部收录在N.David Mermin的Boojums All The Way Through（Cambridge, Eng.：Cambridge University Press，1990）。如果有人对有关这些想法的技术问题感兴趣，那么最好从贝尔本人的论文开始，它们大部分都收录在J.S.贝尔，《量子力学可以言说的和不可以言说的》（Cambridge, Eng.：Cambridge University Press，1997）。

14.虽然定域性假设对爱因斯坦、波多斯基和罗森的论证是非常

重要的，但研究者们努力寻找的却是其论证中其他一些元素的错误，以力图避免得出宇宙允许非定域性存在的结论。比如说，有些人偶尔会声称所有数据需要的是我们放弃所谓的实在主义——物体拥有的被测到的那些性质与测量过程无关。在这里，这样的看法错过了重点。如果EPR的论证得到了实验的证实，那么量子力学的长程关联性也没有什么神秘之处；它们不会比传统的长程关联性（比如你在这边发现左手手套就一定能够在那边发现右手手套）更令人惊奇。但这种推理受到了贝尔埃斯拜科特实验结果的驳斥。现在，即使我们为了回应EPR的驳斥而放弃实在主义——就像我们在标准量子力学中所做的——这也不会对削弱空间上相隔很远的随机过程之间的长程相关性的奇异性有多大帮助；当我们放弃了现实主义，手套就如注释4中所说的那样，成为“量子手套”。放弃现实主义无论如何也不会使观察到的非定域关联看起来不那么诡异。事实上，正是受EPR、贝尔和埃斯拜科特结果的启发，我们试图才保持实在主义——举个例子来说，正如在本章后面部分将要讨论到的玻姆理论中的情形——我们需要用来与数据保持一致的非定域性看起来更加严重，它与非定域的相互作用有关，而不仅仅是非定域关联。许多物理学家反对这样的选项，因此放弃了现实主义。

15.见穆雷·盖尔曼，The Quark and the Jaguar（夸克与美洲豹）（New York：Freeman，1994），以及Huw Price的Time's Arrow and Archimedes' Point（Oxford：Oxford University Press，1996）。

16.狭义相对论不允许任何曾比光速慢的物体超越光速限制。但如果某物总是比光跑得更快，它也不会被狭义相对论严格排除在外。这种假设的粒子被称作超光速粒子。大多数物理学家认为超光速粒子不存在，但有一些人则认为存在这种可能性。根据狭义相对论的方程，这种比光速还快的粒子具有一些特殊性质，所以目前，没有人发现它们有何种特殊用途——即使只是在理论假设上也没什么用。在现

代研究中，人们普遍认为有超光速粒子的理论并不可靠。

17.数学功底不错的读者应该记着，本质上讲，狭义相对论要求物理定律具有洛伦兹不变性，也就是说，物理定律要在闵可夫斯基时空中的SO（3，1）坐标变换下具有不变性。那么，如果量子力学可以完全用洛伦兹不变性方式表达的话，我们就可以说它与狭义相对论相符。现在，相对论性的量子力学与相对论性的量子场论正在朝这个目标努力，但关于它们是否可以在洛伦兹不变的框架内解决量子力学的测量问题，人们还没有达成一致意见。举个例子来说，在相对论性场论中，可直接以洛伦兹不变的方式来计算各种实验结果的概率幅和概率。但标准做法描述了在量子概率范围内的或这或那的特定结果出现的方式——也就是说，在测量过程中发生的事情。这对量子纠缠而言是一个特别重要的问题，因为该现象会受实验者行为——测量纠缠态粒子特性的行为——的影响。有关更深入的讨论，请参考Tim Maudlin, Quantum Non-locality and Relatiuity（Oxford：Blackwell，2002）。

18.对于数学功底不错的读者，下面就是所做预言与这些实验相符的量子力学计算。假设探测器测量到的自旋所绕的轴分别是垂直、绕垂直轴顺时针旋转120度以及绕垂直轴逆时针转120度（就好比有表盘上的中午12点、4点和8点；有两个表盘，每个对应一个探测器、相对放置），为了便于讨论，想象一下，有两个电子，以所谓的单态形式背靠背的朝着探测器飞去。这种总自旋为零的态保证了如果一个电子处于自旋绕某轴向上的态的话，另一个电子将处于自旋向下的态，反之亦然。（还记得吗？在正文中为了方便，我曾将电子之间的关联描述为如果一个电子自旋向上，另一个电子自旋将也向上，如果一个自旋向下，另一个也向下；事实上，关联是指自旋处于相反的方向。为了与正文呼应，你可以将两个探测器想象为刻度相反，因此一个自旋向上，另一个则自旋向下。）基础量子力学的标准结果表明，如果两个探测器所测得的电子自旋所绕的轴之间的角度为θ，那么他

们测量到相反自旋值的概率为$\cos^2(\theta/2)$。因此，如果校准探测器的轴（$\theta=0$），它们一定会测得相反的自旋值（类似于我们在正文中所讲的，当为探测器设定相同的方向时，测得的值总是相同的），如果把轴的方向调成＋120°或-120°，则两个探测器测得相反自旋的概率为$\cos^2$（＋120°或-120°）＝1/4。现在，如果探测器的轴的方向是随机选择的，那么有1/3的概率指向相同的方向，2/3的概率指向相反的方向。因此，整体来看，发现相反的自旋的概率为（1/3）（1）＋（2/3）（1/4）＝1/2，和实验数据一致。

你可能觉得有点奇怪：在定域性假设的前提下，自旋关联的概率（大于50%）竟会大于我们用标准量子力学（恰好50%）算出的自旋关联；你可能会想，量子力学的长程纠缠应该产生更多的关联。事实上，确实如此。这里应当这样想：全部测量中有50%的关联性时，对于左右探测器的轴选定相同方向的测量来说，量子力学将带来100%的关联性。在爱因斯坦、波多斯基和罗森的定域性宇宙中，对于选定相同的轴方向的测量来说，要想有100%的符合，就需要全部测量中有大于55%的关联性。粗略地讲，在定域宇宙中，全部测量中50%的关联性将使得在选择相同轴的前提下出现小于100%的关联性——也就是说比我们在非定域宇宙中发现的关联性要小。

19.你可能会想，从一开始起，波函数的瞬间塌缩就与光速设定的速度上限相矛盾，因而必会与狭义相对论相矛盾。如果概率波像水波那样，你的结论将不容置疑。概率波的值在一个很大的范围内突然衰减为零，简直比太平洋的海面瞬间宁静、海水停止流动还要让人惊奇。但是，量子力学的支持者们提出，概率波不同于水波。虽然概率波描述物质，但它本身并不是物质。这些支持者们继续说，光速限制只适用于物质性对象，而其运动都可以直接看到、感觉到、探测到。如果仙女座星系中电子的概率波衰减为零，仙女座星系的物理学家们将100%的探测不到电子。仙女座星系的任何观测都不能表明概率波的突变与一些成功的探测行为——比如说在纽约发现电子——有关。

只要电子本身没有以大于光速的速度从一个地方运动到另一个地方，就不会与狭义相对论相矛盾。正如你所看到的，唯一发生的事情是在纽约——而不是在别的什么地方——探测到电子。我们甚至都不需要讨论它的速度。因此，当概率波的瞬间塌缩是一种令人困惑不解的迷惑而且有问题的理论（在第7章中有更全面的讨论），但这并不意味着它与狭义相对论矛盾。

20.关于这些观点的讨论，可以参考Tim Maudlin, Quantum Nonlocality and Relatiuity。

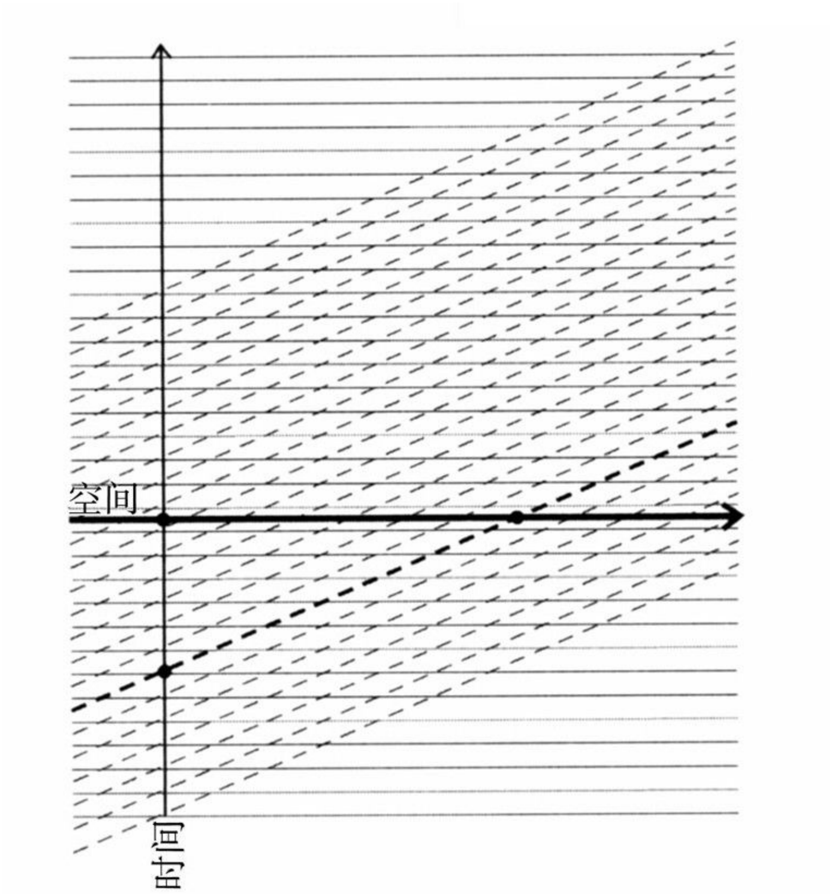
第5章

1.对于数学功底不错的读者而言，从方程$t_{运动}=\gamma[t_{静止}-(u/c^2)x_{静止}]$（在第3章的注释9中曾讨论过）中我们发现丘巴卡在某一给定时刻的现在目录中将包含地球上的观测者声称是$(u/c^2)x_{地球}$之前发生的事情，其中$x_{地球}$表示丘巴卡距离地球的距离。以上的描述是假设丘巴卡远离地球而去。如果是朝向地球的运动，u的方向相反，因此向地球移动的观测者将声称这样的事件发生在$(u/c^2)x_{地球}$之后。令$u=10$英里/小时，$x_{地球}=10^{10}$光年，我们发现$(u/c^2)x_{地球}$大约为150年。

2.这个数字——和后文中描述丘巴卡朝向地球的运动时用到的一个类似的数字——在本书出版时都是有效的。但随着地球上时间的流逝，它们会变得没那么精确。

3.数学功底不错的读者应该注意到，从不同角度切割时空条的比喻正是狭义相对论课程上教授过的时空图概念。在时空图中，从被认为是静止的观测者的角度看，某一特定时刻的所有三维空间都可以用一条水平线来表示（或者，在更准确的图示中，用一个水平面来表示），而时间可以用垂直轴来表示（在我们的叙述中，每个“面包片”——平面——代表的是某一时刻的所有空间，而横穿面包的轴则

是时间轴）。时空图为讲清你和丘巴卡的现在片提供了一种有力的工具。




图中的细实线为相对于地球（为了使问题简化，我们假定地球既没有旋转也没有加速，这些因素对我们要讨论的问题并不重要，只能带来没有必要的复杂性）静止的观测者的等时线（现在片），细虚线为相对于地球以约9.3英里/小时的速度运动的观测者的等时线。当丘巴卡相对于地球静止时，细实线就代表了他的现在片（而你在整个故事中都安静地待在地球上，这些细实线总代表着你的现在片），粗实线代表的是包含了待在地球上的21世纪的你（左边的黑点）和他（右

边的黑点）的现在片，你们俩都老老实实地坐在那里读书。当丘巴卡远离地球而去时，虚线就代表了他的现在片，粗虚线代表的是包含了丘巴卡（正站起来开始走）和约翰·维尔克斯·布思（左下的黑点）的现在片。注意，接下来的一条虚线时间片将包含丘巴卡走动（假设他仍然在四处逛）和21世纪的你静坐着读书。因此，你的某个时刻将出现在丘巴卡的两张现在目录上——一张是他走动之前的，另一张是他走动之后的。这就表明，简单的直觉上的现在概念——在被应用于整个空间时——可以被狭义相对论转换为具有不寻常性质的概念。而且，这些现在名单目录并不会违反因果关系：标准的因果关系（参见第3章注释11）仍然强而有力。丘巴卡的现在目录之所以会突变是因为他自己突然改变运动状态，从一个参考系跃入另一个参考系。但是，对于到底是哪个事件影响了哪个事件，所有的观测者——都用单独一种良好定义的时空坐标——都会认同同一个说法。

4.专业读者会意识到我把时空假设为闵可夫斯基时空。其他几何学中类似的论证不一定会适用于整个时空。

5.《阿尔伯特·爱因斯坦和麦·贝索书信集：1903-1955》（Albert Einstein and Michele Besso Correspondence 1903-1955，P.Speziali编辑Paris：Hermann，1972）。

6.这里的讨论意在对下面的问题给出一个定性上的认识。这个问题就是，你对自己所经历过的生活——正是这种生活经历留给了你那些记忆——有一种感受，这种感受的根基是由你此刻的感受和前一刻的回忆共同形成的，但是这个形成过程是怎样的？比如说，如果你的大脑和身体不知何故达到了与此刻一样的状态，那么你将体验到与你的记忆能够证实的那段经历相同的感觉（假设所有体验的根据都可以在大脑和身体的物理状态中找到），即便那些经历从未发生过，只是人为的印在你的大脑里也没关系。这里的讨论有一个简化之处，即假定了我们能感觉或体验到某一瞬间发生的事情，而事实上大脑识别和诠释它所收到的任何刺激都需要一段时间。不过，这一点虽然不错，

但却与我所要讨论的内容没有特别的相关性；这虽然有趣但却会带来不必要的复杂性，而这种复杂性来自于用与人类的体验直接相关的方式分析时间。就像我们先前讨论到的，以人类为例子可以使我们的讨论更为通俗直观，但我们需要剔除那些从生理上看更加有趣而不是在物理上重要的部分。

7.你可能想知道本章的讨论与第3章中讲过的物体以光速“穿越”时空有什么联系。对那些数学功底不错的读者，这个问题可以粗略的回答为一个物体的历史可以表示为时空中的一条曲线——时空条中的一条路径，这条路径就是粒子所有时刻所在位置的集合（正如我们在图5.1中看到的那样）。穿越时空“运动”这一概念可以直观的表示成指明该路径（但是不要想象成追溯映入眼帘之前的路径）。在这条路上的“速度”可以这样得出：用路径的长度（两点之间的距离）除以该路径上某人或某物携带的表所记录的时间差。我们又一次遇到了不与时间流逝有关的概念：你需要做的只是看一下表在这两点的示数。你会发现，不论运动方式怎样，用这种方式得出的速度都等于光速。数学功底不错的读者会意识到这里的原因是：在闵可夫斯基时空的度规为$ds^2=c^2dt^2-dx^2$（其中$dx^2$是欧几里得长度$dx^2{}_1+dx^2{}_2+dx^2{}_3$），而钟表的时间（“固有”时间）为$d\tau^2=ds^2/c^2$。所以，很明显，刚才所定义的穿越时空的速度可由$ds/d\tau$给出，结果等于c。

8.鲁道夫·卡那夫，“自传”收录在The Philosophy of Rudolf Carnap, P.A.Schilpp编辑（Chicago：Library of Living Philosophers，1963），p.37。

第6章

1.注意，我们提到的不对称性——时间之箭——来自于事件在时间中的发生顺序。你可能也想知道时间本身的不对称性——比如说，就像在后面的章节中我们将看到的，根据某些宇宙理论，时间可能曾有一个开始但可能不会有结束。这些是全然不同的时间不对称性概

念，我们在这儿的讨论主要关注前者。虽然如此，在本章末我们将看到，事件在时间上的不对称性取决于早期宇宙历史的特殊条件，因此，可以将时间之箭与宇宙学的方方面面联系起来。

2.对于有数学功底的读者而言，让我来更为准确地说明一下时间反演对称性意味着什么，并谈谈一个有趣的例外，它对于我们本章中所讨论问题的意义还不完全明了。时间反演对称性概念可以简单地表述为，当一套物理定律的方程有解时，比如解为s（t），s（-t）也是该方程的一个解，则我们称该物理定律具有时间反演对称性。举个例子来说，在牛顿力学中，力取决于粒子的位置，若x（t）＝[$x_1$（t），$x_2$（t），……，$x_{3n}$（t）]是n个粒子在三维空间中的位置，则x（t）为方程$d^2$x（t）/$dt^2$＝F[x（t）]的解，这也就意味着x（-t）也满足于牛顿方程，也就是说$d^2$x（-t）/$dt^2$＝F[x（-t）]。注意x（-t）代表穿越相同位置——比如x（t）——的粒子运动，只不过是以相反的顺序，相反的速度而已。

在更普遍的情况下，物理定律也为我们提供了将物理系统的状态从初始时刻$t_0$演化到t＋$t_0$时刻的运算法则。具体一点说，这种运算法则可以看成映射U（t），它可以将s（$t_0$）映射到s（t＋$t_0$），即：s（t＋$t_0$）＝U（t）s（$t_0$）。如果映射T满足U（-t）＝$T^{-1}$U（t）T，我们就可以说得出U（t）的定律具有时间反演对称性。用通俗的语言来说，这个方程说的是，通过适当运算（用T来表示）物理系统某一时刻的状态，物理系统根据理论定律顺着时间方向在一段时间t内的演化[用U（t）来表示]等于系统逆着时间方向在同样时间t内的演化[用U（-t）来表示]。举个例子来说，如果我们将粒子系统的状态指定为粒子在某一时刻的位置和速度，那么T就可以在保持所有粒子位置不变的情况下使速度反转。这样的粒子分布顺着时间方向在t时间内的演化等于粒子的原始分布逆着时间方向在t时间内的演化（$T^{-1}$因子消除了速度的反转。因此，到最后，不仅粒子的位置回到t个单位时间以前，它们的速度也将如此）。

某些定律中的T运算比牛顿力学中的T运算还要更加复杂。举个例子来说，如果我们研究带电荷粒子在电磁场中的运动，逆向速度还不足以使方程得出粒子回溯的演化。磁场的方向也需要反转（之所以有这种要求是为了使洛伦兹力公式中的uXB项保持不变。）因而，在这种情况下，T运算包含了这两种变换。除了反转所有粒子的速度外，我们还需要做些其他的事情这一事实对正文中的讨论没有影响。重要的是，粒子在一个方向上的运动与物理定律自洽的话，在另一个方向上的运动也与物理定律自洽。我们需要反转磁场不过是某种巧合，并没有什么特别的重要性。

而在弱核力相互作用中事情就不是这么简单了。弱相互作用需要用特殊的量子场论来描述（将在第9章中简要的讨论一下），一个普适的定理证明（还需要保证量子场论具有定域性，幺正性，洛伦兹不变性——这样的量子场论才是人们感兴趣的），量子场论在电荷共轭运算C（将粒子替换为其反粒子），宇称运算P（将粒子的位置变换到镜像位置），以及时间反演运算T（将时间t换为时间-t）的联合作用下总是具有对称性。所以，我们可以将运算?定义为CPT的乘积，但如果?具有不变性就会要求还有一种CP运算存在，?不应被简单的解释为原来的粒子回溯其路径（因为，粒子的种类也会被这样的?运算改变——粒子变成反粒子——所以逆向返回的并不是原来的粒子）。我们会发现，对于某些实验，我们没法很好的解释。对于某些特殊的粒子（比如K介子以及B介子）来说，它们可以在CPT运算下保持不变，但是在单独的T变换下没法保持不变。1964年，詹姆斯·克洛宁，瓦尔·菲奇（因为这一工作，他们两人获得了1980年诺贝尔物理学奖）及其合作者通过证明K介子破坏CP对称性（这就意味着必然会破坏T对称性，因为需要保持CPT不被破坏）间接的确认了这一点。更加晚近的时候，T对称性的破坏通过CERN的CPLEAR实验以及费米实验室的KTEV实验得以直接确立。简单地讲，这些实验证明，如果有人给你展示一段有这些介子参与的过程的影片，那么你就有办法看出这段影片到底是按正确的时间顺序播放，还是反着播放。换句话说，这些特

殊的粒子可以区分过去和未来。还不清楚的是，这与我们每天所感受到的时间之箭是否存在某种联系。不管怎么说，这些奇特的粒子虽然可以在对撞机实验上瞬间产生出来，但却不是我们熟悉事物的组成粒子。对于包括我在内的很多物理学家来说，这些粒子展示的时间不可反转性看起来并不会对时间之箭的谜题有多大影响，所以我们不打算进一步讨论这些例外。但问题是，其实没人真的知道是不是这样。

3.我有时候发现，对于碎蛋壳真的重新聚合起来形成未破碎的蛋壳这样的理论命题，要接受起来还真是挺难的。但是，自然定律的时间反演对称性，正如在前一条注释中更为详细地探讨过的那样，会确保这样的事情能够发生。微观上，鸡蛋的破碎是一个与组成蛋壳的各种分子有关的物理过程。鸡蛋破碎，蛋壳四散，这是因为鸡蛋被摔碎过程中受到的冲击将分子强行拉开。如果这些分子运动反向发生，那么它们就会重新组合在一起，以先前的形式重新成为蛋壳。

4.为了使我们能够集中注意力，以现代方式思考这些思想，我会将一些有趣的历史略去。玻尔兹曼自己对熵这一主题的思考在19世纪70～80年代经历过重大精炼。在那个时期，他与一些物理学家的相互影响和交流对他来说是非常有帮助的，这些科学家包括詹姆斯·克拉克·麦克斯韦，开尔文勋爵，约瑟夫·洛施密特，约什亚·威拉德·吉布斯，亨利·庞加莱，S.H.勃柏利，以及欧内斯特·切梅罗。事实上，玻尔兹曼最初认为他可以证明，对于孤立的物理系统，熵会一直并且绝对化的不减，而不只是这样的熵减过程不太可能发生。但是来自上述以及其他一些物理学家的反对意见，促使玻尔兹曼强调这个问题中的统计或概率方法，而这种方法一直用到了今天。

5.我想象着我们正在用现代经典文库（Modern Library Classics）版的《战争与和平》，Constance Garnett译，1386页。

6.数学比较好的读者应该会注意到，由于数太大了，熵实际被定义为可能的排列数的对数，不过这一细节在这里与我们无关。但是，

从理论的角度看，这一点非常重要，因为熵是所谓的广延量这一点非常方便，这意味着如果你把两个系统合在一起，那么整体的熵就是两个系统各自的熵的和。而仅当熵是对数的形式这一点才会成立，因为在这种情况下总的排列数会等于各自排列数的乘积，因而排列数的对数是相加性的。

7.尽管在理论上，我们可以预言每一页被放在哪里，你可能还是会关心决定着页序的另一个元素：你怎样把这些页整齐的摞在一起。这与所要讨论的物理无关，但如果它令你心烦的话，你可以这样想象：我们同意你把它们一张张地捡起，从离你最近的那张开始，然后再捡起离那张最近的一张，如此下去（而且，怎样确定最近的纸张都是一样的，比如说，我们都同意从纸张距离我们最近的一个角开始量起）。

8.以为对于不多的一些页，就能在达到预言其页序（运用某些方法将它们堆在一起，参见前一条注释）的精度上成功的算出其运动，实际上是极其乐观的想法。根据纸张柔韧性与重量的不同，这样一个相对“简单”的计算也远远超越了今天的计算机能力。

9.你可能会担心，定义页序的和定义一群分子的熵之间存在着根本性的差异。毕竟，页序是离散的——你可以一页一页的数清它，虽然所有的可能性总数会很大，但却是确定的。而相反的是，即使单独一个分子，其运动和位置都是连续的——你没法一个一个的数清，因而可能性的总数会是无限大（至少在经典物理中是这样）。所以，我们怎样数清分子的排列数呢？这个嘛，简单地说，这是个好问题，但却是个被完全解决了的问题——要是这能让你感到满意，那你就不用看下面的这段内容了。详细的回答你这个问题需要用到一点数学，要是没有基础的话可能不太好理解。物理学家用相空间——一个6N维空间（N为粒子数），其中的每个点代表着一个粒子的位置和速度（描述每个粒子的位置需要3个数，速度也是一样，因而N个粒子需要6N个数）——来描述经典多粒子系统。关键之处在于，相空间可以被划

分为不同区域，给定区域内的所有点对应着具有外在相同整体性质的分子的速度速率分布。如果相空间给定区域的分子排布从一个点变化到同一区域的另一个点，那么在宏观上我们是没法区分这种变化的。现在，我们就不需要数清给定区域内点的数目——与数清不同页序的排列数最最直接的类比，但这种类比会导致无限大——物理学家们就用相空间中每一个区域的体积来定义熵。体积越大，则区域内的点越多，因而熵就越大。而区域的体积，即使是很高维度的空间中的区域，可以有严格的数学定义。（数学上，需要选取所谓的测度，对数学比较好的读者，我要指出，对于与给定宏观态相一致的所有微观态，我们通常选取的测度都是一样的——也就是说，与给定宏观性质有关的微观分布被假定为等权重。）

10.特别是，我们知道一种它可能发生的方式：如果几天前，$CO_2$还在瓶中，那么从我们上面的讨论中我们可以知道，如果你现在同时反转每一个$CO_2$分子的速度，以及每一个与$CO_2$分子有相互作用的分子的速度，等上几天后，你将发现所有的$CO_2$分子又聚集起来回到了瓶中。但是这种速度反转没法应用于实践，每件事看起来只能按其自己的步调发生。不过我需要指出，我们可以在数学上证明，只要等待的时间足够长，$CO_2$分子早晚会按自己的步调退回到瓶中。19世纪，法国数学家约瑟夫·刘维尔证明的一个结果可以用来构建所谓的庞加莱可逆定理（Poincaré recurrence theo-rem）。根据这一定理，如果你的等待时间足够长，一个有限体积内具有有限温度的系统（比如封闭空间内的$CO_2$分子）将会达到与其初始态任意接近的态（在本例中，所有的$CO_2$分子都被封闭在可乐瓶中）。问题在于你到底得等多久它才会发生？对于一个其组分很多的系统来说，这个定理告诉我们，要想使其按自身步调回到其初始状态，你的等待时间可能会比宇宙的年龄还长。然而，理论上重要的是，如果你有足够的耐心，能够等待足够长的时间，那么空间中所有的物理系统都会回复到其初始状态。

11.你可能会想，水为什么会结成冰呢？那岂不意味着$H_2O$分子变

得更加有序，换句话说，岂不是获得了较低而不是较高的熵？这个嘛，大体上说来，液态的水变成固态的冰，会向周围的环境释放能量（而当冰化成水的时候，则会从环境中吸收能量），而这会导致环境中的熵有所提高。当环境温度足够低的时候，低于零摄氏度，环境中的熵增超过了水中的熵减，因此，结冰过程是一个熵减少的过程。这就是冬天为什么会结冰的原因。类似的，当你冰箱中的冰块形成时，$H_2O$分子中的熵固然是减少了，但是在这个过程中冰箱会向周围的环境释放热量，因而会导致总的熵是增加了。对数学比较好的读者，更加准确的说法是，我们所讨论的这种自发现象由所谓的“自由能”掌控。直观上，自由能是一个系统中可以用来做功的那部分能量。数学上，自由能F，由F＝U-Ts来定义，其中U代表总能量，T代表温度，s代表熵。如果某一过程会导致自由能减少，那该过程就可以自发产生。低温下，液态水中U的减少超过了固态冰中s的减少（超过了-Ts的增加），因而这个过程会自然发生。而在高温下（零摄氏度以上），冰由固态到液态或气态的转变则是符合熵的要求的（s的增加超过了U的改变），因而也会自然发生。

12.直接应用熵的论证究竟是怎样使我们得出记忆与历史记录是对过去不可信赖的记述这一结论的。关于这个问题，要想看看较早的讨论，可以参看C.F.von Weizs?cker的The Unity of Nature（New York：Farrar, Straus and Giroux，1980）中138-146页，最初发表在Annalen der Physik 36（1939）。要想看较新些的优秀作品，可以看看大卫·阿尔伯特的Time and Chance（Cambridge, Mass.：Harvard University Press，2000）。

13.事实上，因为物理定律无法区分时间上的前与后，所以对半小时前，晚上10点，完全结成冰块的解释同预言半小时后，晚上10点，小冰渣完全结成冰块一样，都可以说是极为荒唐——从熵的角度讲。相反，对晚上10点时液态的水，到了晚上10点半的时候慢慢形成了冰渣这一现象的解释，则与预测到了晚上11点的时候，那些冰渣又

化为水在道理上是一样的，都是我们熟悉且可预期的事情。对后一种现象的解释，从晚上10点半看，前后的结冰和融化现象完全是对称的，而且与我们的实际观测是相符的。

14.特别细心的读者可能会想到，我在讨论中错误地使用“早期”这个词，因为这相当于插入了时间上的不对称性。按更为准确的语言，我想表达的是，我们会需要特殊的条件使时间维度的一端变得特别。在后面将会更加清楚，特殊的条件就是低熵的边界条件，而我所谓的“过去”，就是时间维度上满足这一条件的那一端。

15.时间之箭要求低熵的过去这一想法已经有很长的历史了，可追溯到玻尔兹曼及其同时代的人；在汉斯·雷肯巴赫著的The Direction of Time中有详细讨论（Mineola, N.Y.：Dover Publication），在罗杰·彭罗斯著的《皇帝的新脑》中以特别有趣的定量方式有所讨论。

16.回想一下，我们在本章中的讨论并没有考虑量子力学。如史蒂芬·霍金于20世纪70年代证明的那样，当量子力学的效应被考虑进来时，黑洞会允许一定数量的辐射逃离出去，但这不会影响他们称为宇宙中最高熵的物体。

17.一个很自然的问题是，我们是怎么知道未来不会有新的对熵有影响的限制出现。答案是我们没办法知道，某些科学家甚至提出一些实验，用以探测这些将来的限制对我们今天能够观测到的事物的影响。有一篇非常有趣的文章探讨了将来的和过去的一些对熵的限制的可能性，即Murray Gell-Mann与James Hartle合著的Time Symmetry and Asymmetry in Quantum Mechanics and Quantum Cosmology，收录在J.J.Halliwell, J.Pérez-Mercader, W.H.Zurek编辑的Physical Origins of Time Asymmetry（Cambridge, Eng.：Cambridge University Press，1996）。同一文集的第四和第五部分中也有一些关于这个问题的有趣文章。

18.在这一章中，我们一直在用“时间之箭”这个词，用以表达时空的时间轴（任意观测者的时间轴）具有不对称性这一明显事实：按时间轴的方向排列顺序的事件非常多，但是反向顺序的事件，即便不是没有，也是很少发生。很多年来，物理学家和哲学家一直在将事件顺序分成不同子类，在这些子类中，至少在理论上，时间上的不对称性可以归结于逻辑上独立的解释。比如说，热量总是从热的物体流向较冷的物体，反之则不行；电磁波总是从恒星或者电灯泡这样的源射出，看起来却永远不会聚拢回这些源；宇宙看起来是在碰撞，而不是在收缩；我们记住的是过去而不是未来（这些分别被称为热力学、电磁学、宇宙学、心理学上的时间之箭）。所有的这些都是时间不对称现象，不过至少在理论上，这些现象有可能从全然不同的物理原理中获得其各自的时间不对称性。我的观点（很多人也有这样的观点，而另一些人则不）是，除宇宙学上的时间之箭外，其他的时间不对称现象在基本层面上没有什么区别，它们都可以被归结为同样的解释——我们在本章中讲过的解释。比方说，电磁辐射为什么向外传播而不是向内传播？要知道这两种情形都是麦克斯韦方程的解。这个嘛，由于我们的宇宙有提供这些向外放射的波的低熵、连贯、有序的源——比如说恒星或电灯泡——而这些有序的源的存在又可追究到宇宙起源时的更为有序的环境，参见正文中的讨论。解释心理学上的时间之箭要稍微难一些，因为对于我们还没有搞懂的人类思想来说，还没有微观物理基础。但是，对于与计算机有关的时间之箭，人们却取得了一些进展。进行计算，完成计算，记录结果，对于这些基本的计算顺序，其中的熵的性质人们已经搞清楚了（查尔斯·本耐特，罗尔夫·兰道尔和其他一些人的贡献），且与热力学第二定律符合得非常好。因而，如果人类思想类似于计算过程，我们就可以对其应用类似的热力学解释。但是，也需要注意到，与宇宙正在膨胀而不是收缩这一事实紧密联系的不对称性，实际上与我们一直在探索的时间之箭有关，但在逻辑上却是完全不同的。即使宇宙慢慢减速，停止下来，再开始转入收缩过程，时间之箭还是会指向同一个方向。即便宇宙膨胀过程反转过来变成了收缩，物理过程（鸡蛋破碎，我们变老，等等）还是会像往

常一样发生。

19.数学比较好的读者应当注意到，当我们使用这种概率声明时，我们假定了一种特别的概率测度，从而，与我们此刻所看到的宏观事物相容的所有微观态都有相同的测度。当然，还有另外一些我们能用的测度，比如说，大卫·阿尔伯特在Time and Chance中倡导使用的概率测度，对于所有的微观态——与我们此刻所看到的宏观事物以及他所谓的过去假设（the Past hypothesis，宇宙开始于低熵态这一明显的事实）相容的微观态——都是一样的。利用这样的测度，我们可以不用考虑那些与低熵的过去——为我们的记忆、记录以及宇宙学理论所确证——不相容的历史。按这种思考方式，一个低熵的宇宙是没有概率问题的；根据假设，宇宙按低熵方式开始的概率为100%。但仍有一个重大问题：宇宙为什么会按那种方式开始？关于宇宙的问题甚至都没在概率背景下表述。

20.你可能会忍不住争辩，已知宇宙在早期之所以低熵不过是因为其尺寸远比今天要小，因而，正如书页较少的书熵也低些，早期宇宙中的组分所可能有的排列数要更少些。但是，只依靠这种说法说服自己是没有用的。很小的宇宙也可以有很高的熵。比方说，我们宇宙的命运之一（尽管很不可能）就是在未来的某一天停止膨胀，然后向内爆裂，结束于所谓的大收缩（big crunch）。计算表明，即使宇宙的尺寸在向内爆裂阶段减小，熵也仍会增加，这就显示了小体积并不一定意味着低熵。不过，我们将会在第11章中看到，宇宙初期的小尺寸的确会在我们当今对低熵起源的最佳解释中占有一席之地。

第7章

1.众所周知，如果你所研究的是三体乃至更多体相互作用的运动问题，经典物理的方程是不能精确求解的。所以，即使在经典物理中，任何有关大群粒子运动的语言也都只能是近似的。但问题在于，在经典物理中，对近似的精确度不会有基本层面的限制存在。在经典

物理掌控的世界中，所用的计算机能力越强，位置和速度的初始数据给的越精确，我们所能得到的结果就越精确。

2.在第4章结尾，我注意到贝尔、艾斯柏克特以及其他人的结果并没有排除粒子总是有确定的位置和速度——即便我们没法同时确定这些量——这种可能性。而且，玻姆版的量子力学清楚地实现了这种可能性。因而，尽管人们普遍认为，像电子在被测量之前不会有一个位置这种事，属于传统方法的量子力学中的标准性质。但是严格说来，将其作为一条普遍适用的声明未免太牵强了些。记住，在玻姆的方法中，我们将会在本章稍后讨论到，粒子总是与概率波“相伴”；也就是说，玻姆的理论总要用到粒子与波，而标准方法用的是互补性，可概括为粒子或波。因而，我们所探求的结论——如果我们单说一个粒子在每一个确定的时刻通过空间中一个确定的点（在经典物理中我们就会这么做），那对过去的量子力学描述绝对是不完备的——但还是正确的。在传统方法的量子力学中，我们必须将一个粒子在给定时刻所能占据的大量位置都考虑进去，而在玻姆的方法中，我们也必须将“导”（Pilot）波——而它也可以延展至很多个位置——包括进去（专家级读者应该会注意到，导波只不过是传统量子力学中的波函数，尽管其在玻姆理论中的化身相当不同）。为了避免无边的限定，我们在接下来的讨论中将使用传统的量子力学观点（最为广泛使用的方法），对玻姆以及其他方法的评论将放在本章最后。

3.要想看看数学化一点但同时又高度教学式的说明，可以参考理查德·费曼与A.R.希布斯合著的Quantum Mechanics and Path Integrals（Burr Ridge, Ill.：McGraw-Hill Higher Education，1965）。

4.你可能会忍不住援引第3章中的讨论——从那里的讨论中我们了解到在光速时时间停止——来辩称，从光子的角度看，所有的时间都一样，所以，光子在经过分束器时“知道”探测器的开关如何设定。但是，我们用其他慢于光速的粒子，比如光子，也可以做这样的实

验，而结果不变。因而，这样的角度不能揭示物理实质。

5.这里讨论的实验设置以及实际确认的实验结果，来自于Y.Kim, R.Yu, S.Kulik, Y.Shih, M.Scully, Phys.Reu.Lett, vol.84，No.1，pp.1-5。

6.量子力学也可以基于一个沃纳·海森堡于1925年提出来的不同形式（所谓的矩阵力学）的等价方程。对数学比较好的读者，薛定谔方程可以写为：Hψ（x, t）＝ih[dψ（x, t）/dt]，其中H代表哈密顿量，ψ代表波函数，而h为普朗克常数。

7.专业读者应该会注意到我在这里略掉了不太明显的一点。即，我们必须得对粒子波函数取复共轭，以保证其为时间翻转版的薛定谔方程的解。也就是说，第2章注释6中讲到的T操作将波函数ψ（x, t）映射为ψ*（x，-t）。这对文中的讨论没有影响。

8.玻姆实际上重新发现并进一步发展了一种可追溯到路易·德布罗意公爵的方法，所以，这种方法有时候被称为德布罗意玻姆方法。

9.对数学比较好的读者来说，玻姆的方法是在位型空间中定域的但在真实空间中非定域的。波函数在真实空间中一个位置处的改变会立即对位于远处的其他粒子产生影响。

10.要想看看对Ghirardi-Rimini-Weber方法极为清楚的处理以及其在理解量子纠缠中的作用，可以看看J.S.贝尔的“Are There Quantum Jumps？”收录于Speakable and Unspeakable in Quantum Mechanics（Cambridge, Eng.：Cambridge University Press，1993）。

11.有些物理学家认为，单子上所列的问题不过是量子力学早期认识不清所带来的无关副产品。按这种观点，波函数只不过是用来预言（概率）的理论工具，除了数学上的实在性不应再有任何其他实在

性（这种方法有时候被称为“闭嘴，去计算”方法，因为这种方法鼓励人们用量子力学和波函数的语言，而不要费力气去想波函数的实际意义）。而另一种不同的方案则认为波函数永远也不会真的塌缩，只不过是其与环境的相互作用使它看起来像是塌缩了（我们稍后将讨论这种方法的一个版本）。我很赞同这种想法，并且实际上，我真的相信波函数塌缩的概念最终将被废弃。但是，我对前一种方法不太满意，就像我不准备放弃思考这个问题：当我们“不看”这个世界的时候它到底发生了些什么。至于后一种方法——按我的看法，在正确的方向上的那个——还需要进一步的数学发展。最后应该得到的结果是，测量带来的某些东西，就是，或者类似于，或者看起来像是，波函数塌缩。要么通过更好的理解外界影响，要么通过尚未发现的其他方法，必须搞清楚波函数到底是怎么回事，而不能简单的扔到一边不加理会。

12.还有另外一些与多世界诠释有关的争议性问题超出了多世界理论本身的夸夸其谈。比如说，在每一个观测者都有无限多个复制品的框架下，定义概率的概念本身就是一种挑战，因为对所谓的观测者来说，其测量被假定为归结于那些概率。如果某一给定的观测者只是众多复制品中的一个，那么当我们说他或她以某一特定的概率测得或这或那的结果时，我们表达的究竟是什么意思呢？哪一个才是真正的“他”或“她”？这个观测者的每一个复制品都将测得——以100%的概率——他或她所存在于其中的那个宇宙或这或那的结果，所以，在多世界体系中，整个概率体系都需要仔细探究。而且，数学比较好的读者应该会注意到一个技术性的问题，根据在如何精确定义多世界上的不同，人们需要选出特定的本征基矢。但是，本征基矢究竟该怎么选出呢？关于所有的这些问题，人们有很多讨论及文章。但是到目前为止，还没有哪个答案被普遍接受。基于我们随后将讨论到的退相干概念的方法对这些问题会有所帮助，特别会对本征基矢的选择问题有所启示。

13.玻姆或德布罗意玻姆方法从未取得过广泛关注。或许原因之一在于，如贝尔在他的文章“The Impossible Pilot Wave”（收录在Speakable and Unspeakable in Quantum Mechanics中）中指出的那样，就连德布罗意和玻姆对他们发展的那套理论都没有特别的好感。但是，也正如德布罗意指出的那样，德布罗意玻姆方法扫清了更为标准的方法中的很多含混不清以及主观性的东西。如果没有其他理由，即使这个方法是错的，能够知道下面这个事情也是值得的，即，粒子在所有时刻都可以有确定的位置和确定的速度（即使在理论上，这也是我们力所不及的测量），且标准的量子力学预言——不确定性及所有其他——全部得以确认。反对玻姆方法的另一条意见是，这个体系中的非定域性比标准量子力学更为“严重”。这就意味着，玻姆方法从最外层起就将非定域相互作用作为理论的核心元素；而在量子力学中，非定域性则埋藏得很深，只能通过间隔很远测量中的非定域关联才能得以体现。但是，正如这一方法的支持者论辩的那样，隐藏起来并不代表会少出现，而且，标准量子力学在测量问题——只有在这个问题上非定域性才会显现出来——上含混不清，而一旦这个问题得以解决，那么非定域性可能就不会再藏在暗处了。还有一些人提出，创建玻姆理论的相对论版本时会遇到很多障碍，虽然在这方面也取得了一些进展（可以参见，比方说贝尔的Beables for Quantum Field Theory，收录在上面提过的同一文集）。所以，在脑海中记住有另一种方法存在无疑是有益的，即便最后量子力学证明另外的方法不过是螳臂当车也没关系。数学比较好的读者可以看看Tim Maudlin的《量子非定域性与相对论》（Malden, Mass.：Blackwell，2002），其中对玻姆理论以及量子纠缠的论述非常不错。14.有关更普遍意义下的时间之箭更为深入以及技术性的讨论，特别是退相干所起的作用，可以参考H.D.Zeh著的The Physical Basis of the Direction of Time（Heidel-berg：Springer，2001）。

15.下面给出的例子将有助于你了解退相干发生得到底有多快——环境的影响将以多快的速度压过量子干涉，从而将量子概率转变

为熟悉的经典物理。下面就是一些例子，其中涉及的数字只是个大概，但所要传达的要点却是很清楚的。在你屋子里漂浮的灰尘，其波函数在空气分子涨落的影响下，会在万亿亿亿亿分之一（10-36）秒的时间内退相干。如果这粒尘埃被完美的隔离，只能与阳光有相互作用，那么其波函数的退相干就会慢一些，但也只有十万亿亿分之一（10-21）秒。如果这粒尘埃漂浮在漆黑一片、空无一物的外太空中，且只与大爆炸之后残留下来的微波背景光子相互作用，那么其波函数将会在百万分之一秒的时间里退相干。这些数字都极小，这意味着即使尘埃这么小的东西的退相干过程都发生得如此之快。而更大的物体，退相干过程发生得还要更快。毫无疑问，即便我们的宇宙是量子宇宙，我们周围的世界看起来也只能就是现在这副模样（可以参考，比如说，E.Joos的“Elements of Environmental Decoherence”，收录在Decoherence：Theoretical, Experimental, and Conceptual Problems, Ph.Blanchard, D.Giulini, E.Joos, C.Kiefer, I.-O.Stamatescu等编辑[Berlin：Springer，2000]）。

第8章

1.更准确地说，康涅狄格的物理定律与纽约的物理定律之间的对称性不仅仅有平移对称性还有转动对称性。当你在纽约表演你的体操套路时，相对于你在康涅狄格的练习，你所改变的不只是位置，你所面对的方向很可能也发生了改变（比方说练习的时候朝东而比赛的时候朝北）。

2.通常情况下，我们会说牛顿运动定律相应于“惯性观测者”；但是如果我们进一步追究究竟何为惯性观测者，则我们将会陷入概念循环：惯性观测者指的是牛顿定律对其完全成立的观测者。要想搞明白这究竟是怎么回事，可以这样想：牛顿定律只是将我们的注意力都转移到一大类特别有用的观测者身上了，而这类观测者对运动的描述完全的并且定量的符合牛顿理论的框架。根据定义，这些就是惯性观测者。技术上讲，所谓的惯性观测者就是那些不受任何力作用的观测

者，亦即那些不处于加速状态的观测者。而爱因斯坦的广义相对论则与之相反，广义相对论可应用于所有的观测者，无论其是否处于运动状态。

3.如果在某段时期，我们周围所有的变化都停止了，我们就感受不到时间的流逝了（身体及大脑的所有功能当然也不运作了）。但这是否就意味着图5.1中的时空条走到了尽头呢？或者换种说法，是否在时间轴上不再有变化——也就是说，时间能不能走到尽头或者以某种形式上的意义继续存在——是一个既难以回答又与我们的生活体验和感受相去太远的假想问题。需要注意的是这一假想情形与熵不能继续增加的最大无序态是有区别的，在那个问题上，虽然熵不能增加，但是微观上的改变，比如气体分子四散流动，仍然可以发生。

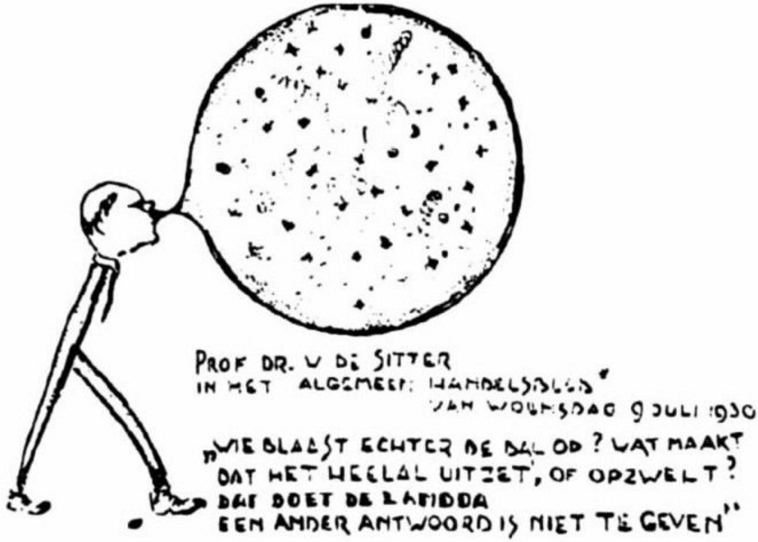
4.宇宙微波背景辐射于1964年由贝尔实验室的两位科学家阿诺·彭齐亚斯与罗伯特·威尔逊在测试卫星通信上使用的大型天线时首先发现。彭齐亚斯和威尔逊遇到了无法消除的背景噪声（他们甚至对掉落在天线内的鸟粪——“白噪声”——也不放过）；在普林斯顿的罗伯特·迪克及其学生皮特·洛尔、大卫·威尔金森与吉姆·皮伯尔斯的关键性理解的帮助下，他们最终认识到天线中出现的噪声实际上是起源于宇宙大爆炸的微波辐射（乔治·伽莫夫、拉尔夫·奥弗尔和罗伯特·赫尔曼在此前即已做出的宇宙学上的重要工作为微波背景辐射的发现奠定了理论基础）。我们在后面的章节中将会论及，微波辐射带给我们的是30万年前的宇宙的清晰图像。在远古时期，干扰光束正常运动的电子与质子这样的带电粒子，通过相互作用形成电中性的原子，从而使得光大体上可以自由传播。从那以后，这些远古时期的光——产生于宇宙早期的光——就不受阻碍的穿行于宇宙之中，到了今天就成了遍布宇宙的微波光子。

5.这里所涉及的物理现象（将在第11章中有所讨论）即是所谓的红移。普通的原子，比如氢原子与氧原子，会发射出一定波长的光，实验室中的实验就能很好地记录下这些光的波长。当这样的物质作为

远去的星系的组分时，它们所发出的光的波长变长了，就像远去的警笛声被拉长音调变低了一样。而红色波长的光是肉眼所能看到的最长波长的光，所以我们把这种光的波长的拉伸效应称为红移效应。红移效应随着远离速度的变大而增加，因而只要将所测得的光波长与实验室中的光波长相比较，我们就可以算出远方天体的速度（实际上这只是红移中的一种，即所谓的多普勒效应。引力也可以引起红移现象：光子在逃出引力场的时候其波长将变大）。

6.更准确地说，热衷于数学的读者会注意到，若一个半径为R而质量密度为ρ的球体表面有一个质量为m的粒子，则该粒子将处于加速状态，$d^2R/dt^2$为（4π/3）$R^3G\rho/R^2$，故有（1/R）$d^2R/dt^2$＝（4π/3）Gρ。如果我们简单地将R视作宇宙半径，ρ为宇宙密度，那么这就是关于宇宙演化尺寸的爱因斯坦方程（忽略压强的影响）。

7.参见P.J.E.Peebles, Principles of Physical Cosmology（普林斯顿：普林斯顿大学出版社，1993），81页。图示中写道：“谁真的能够吹破这个球呢？究竟是什么使宇宙膨胀？是Lambda！除此之外再也给不出其他答案了。”（Koenraad Schalm翻译。）Lambda即所谓的宇宙常数，在第10章中我们将对其有所探讨。




8.为避免混淆，我们需要注意到硬币模型的一个缺点是所有的硬币都彼此相同，而对于星系则显然并非如此。不过关键在于在最大的尺度上——一亿光年的距离上——星系之间的个体差异可以被平均掉，因而当我们分析的是非常巨大的空间体时，每一个这样的空间的整体性质上的差异可以被忽略不计。

9.你也可以小心翼翼地飞到黑洞外缘上，停留在那里，始终开着飞船引擎，以防被黑洞吸进去。黑洞奇强无比的引力场会使时空强烈弯曲，而这会使你的手表相比于你在星系中其他普通位置（相对空旷很多的宇宙空间）时慢得多。你的手表记录的时间依然有效。但是，就像你开着快车到处转时对时间的感受并不总是符合实际时间一样，你在黑洞外记录的时间也只是你的个人体验。当我们要把整个宇宙当做一个整体研究时，对我们更有益的是能够广泛应用与接受的时间概念，而这样的时间概念需要由沿着宇宙空间膨胀方向运动，属于更加微弱更加均匀的引力场中的钟来提供。

10.数学较好的读者应当注意到光将沿着时空度规的类光测地

线，为明确起见，我们取$ds^2=dt^2-a^2(t)(dx^2)$，其中$dx^2=dx^2_1+dx^2_2+dx^2_3$，$x_i$为随动坐标。令$ds^2=0$以满足类光测地线的要求。我们将在时间t发射的光传播到时间$t_0$时的总随动距离写作$\int^{t_0}_t[dt/da(t)]$。如果我们用时间$t_0$时标度因子$a(t_0)$乘以该积分的值，我们就算出了在此时间间隔内光所传播的物理距离。这种算法被广泛用于计算给定时间间隔内光的传播距离，以便了解空间中的某些点之间，比如任意两点之间，是否有因果性联系。如你所见，对于加速膨胀来说，即便是任意大的$t_0$，该积分的值也是有限的，而这表明光并不能传播到任意远的随动位置。因而，在一个加速膨胀的宇宙中，存在着我们永远无法与之联系的区域，或者反过来说，存在着永远无法与我们取得联系的区域。我们称这些区域处于我们的视界之外。

11.当分析几何形状的时候，数学家和物理学家会采用一种定量的曲率方法，这种方法发展于19世纪，是今天所谓的微分几何的一部分。关于曲率的测量，我们可以用一种技术性不那么强的思考方式，即考虑画在感兴趣的形状上的三角形。如果这个三角形的所有内角加起来等于180度，就像画在平面上的三角形那样，我们就说这一形状是平直的。但如果该三角形的内角和加起来大于或小于180度的话，例如画在球面上（因为球面向外膨起，因而画在球面上的三角形内角和大于180度）或马鞍面上（马鞍面则是向内收缩，从而使得画于其上的三角形内角和小于180度）的三角形，我们就说这个形状是弯曲的。如图8.6所示。

12.要是你将一个圆环面的两个不同面的垂直边缘粘在一起（之所以可以这样是因为两个面是可以区分的——你在一个面上走到尽头就到了另一面）你就得到了一个柱面。接着，你再同样地把上下边缘（现在这个上下边缘是一个环的上下边缘了）粘在一起，你就得到了一个油炸面圈。因而，油炸面圈是一种想象或者表示圆环面的方式。这种表示的复杂性在于油炸面圈不再像圆环面一样扁平了！但它实际

上还是扁平的。利用前一条注释中给出的曲率概念，你会发现画在油炸面圈表面的各种各样三角形的内角和都是180度。油炸面圈看起来鼓鼓囊囊的完全是我们将一个两维形状嵌入到三维世界所造成的假象。出于这样的理由，我们在当前章节的内容中使用两维或三维环面明显不弯曲的表示。

13.需要注意我们已经放宽了对形状和弯曲概念的区别。完全对称的空间一共有3种曲率：正，零，负。两种形状可以完全不同但却有相同的曲率。比方说，平面显示屏和无限大的平直桌面。因而，对称性可以帮助我们将空间的弯曲减少为3种可能，但是这3种不同弯曲方式的空间形状则可能有较多种（根据数学家们所关注的整体性质来加以区分）。

14.到目前为止，我们所关注的只是三维空间的曲率——时空片中空间部分的曲率。但是，虽然很难给出图像，我们还是需要知道，对于所有的3种空间曲率（正、零、负），整个的四维时空都是弯曲的，我们研究的宇宙越接近于大爆炸，其曲率就越大。事实上，对于大爆炸时刻附近的宇宙，时空的四维曲率变得如此之大以至于爱因斯坦方程不再成立。我们将在后面的章节中详细的探讨这一问题。

第9章

1.如果你将温度提升到更高，你会发现物质的第四态，所谓的等离子体态。当物质处于这种状态时，物质中的原子进一步分解为组分粒子。

2.有一些很奇怪的物质，比如说罗谢尔盐，这些物质会在高温下变得无序，而在低温下变得有序——与我们通常认为的变化反着来。

3.物质场与力场的一大区别可以表述为沃尔夫冈·泡利不相容原理。这一原理指出，尽管大量的力的粒子（比如光子）可以组合在一

起成为量子时代之前的物理学家（如麦克斯韦）可接受的场，你每次走进黑暗的房间打开电灯时看到的场；但是，根据量子物理的定律，物质粒子却不能按这样连续的、有组织的方式协作（或者更准确地说，两个相同种类的粒子，比如两个电子，不能占据相同的量子态，而对光子则没有这样的限制。因而，物质场不能有这种宏观的、类经典的显现形式）。

4.在量子场论的框架下，所有的已知粒子都被看做是潜在的场的激发态，每种粒子都与相应的场有关。光子是光子场——也就是电磁场——的激发态；上夸克是上夸克场的激发态；电子是电子场的激发态，如此等等。按这种方式，所有的物质与所有的力都可以用统一的量子力学语言加以描述。一个关键问题是，用这种语言描述引力的所有量子性质已被证明是极度困难的，我们将在第12章中讨论这一问题。

5.希格斯场是以彼得·希格斯命名，但还有很多其他物理学家——托马斯·基布尔，菲利普·安德森，R.布劳特，弗朗索瓦·恩格尔特以及其他人——在将希格斯场引入物理学中以及理论发展中扮演过重要角色。

6.记住，场的值决定于其到碗心处的距离，所以，即便其场值落到碗谷时场的能量为零（因为高于碗谷之处代表着场的能量），其场值也并非为零。

7.在正文中，希格斯场的值由其与碗中央的距离给出。你可能会问，碗的圆环形谷底上的点——这些点到碗中央的距离都一样——究竟是怎样给出任意而不是相同的希格斯场值呢？答案是，数学比较好的读者很容易明白，谷底不同的点所表示的希格斯场的值在大小上是一样的，但是其相位却不同（希格斯场的值是复数）。

8.理论上，物理学中的质量概念有两个。其中之一是正文中讲过

的概念：质量是物体用以抗拒加速的性质。有的时候，这样定义的质量被称为惯性质量。第二个质量概念与引力有关：物体的质量表征的是该物体在给定强度的引力场（比如地球的引力场）中所受到的引力的大小。有时，这一质量概念被称为引力质量。乍看之下，希格斯场只与惯性质量有关。但是，广义相对论的等价原理声称，一个物体在加速运动中感受到的力与在相应大小的引力场中感受到的力不可区分——两者完全等价。这就意味着惯性质量与引力质量在概念上的等价。故而，希格斯场与我们提到的两种质量都有关，因为根据爱因斯坦的理论，两者是等价的。

9.我在这里感谢拉斐尔·凯斯帕，他向我指出这里的描述是大卫·米勒教授的获奖比喻的一个变种。1993年，英国科学部长威廉·瓦德格雷夫向英国物理学会提出一个挑战，让物理学家们提出最好的比喻来说明为什么纳税人的钱应该花在寻找希格斯粒子上。大卫·米勒教授赢得了这一挑战。

10.数学比较好的读者应该认识到，光子、W玻色子以及Z玻色子在电弱理论中同属于SU（2）XU（1）群的伴随表示，因而在该群的作用下可以交换。而且，电弱理论的方程在这个群的作用下会展现出完全的对称性，在这种意义上，我们说力的粒子彼此关联。更为准确地说，在电弱理论中，光子是明显的U（1）对称性的规范玻色子与SU（2）的U（1）子群的规范玻色子的一种特殊混合；因而，光子与弱规范玻色子有很密切的关系。但是，由于对称群的乘积结构，4种玻色子（实际上有两种W玻色子，彼此电荷相反）在群的作用下并非完全混合。在某种意义上，弱相互作用与电磁相互作用是一种数学体系的不同部分，但这种数学体系本身并不是完全统一的。当强相互作用被包括进来的时候，这个群变得更大，加入了一个SU（3）因子——“色”SU（3）——这样一来这个群就有3个独立因子SU（3）XSU（2）XU（1），这样进一步表明完整统一性的缺失。这也是将在下面小节中讨论的大统一理论的部分动机：大统一理论试图找到一个

可以描述更高能标上的力的半单（李）群——只有一个因子的群。

11.数学比较好的读者应该认识到，乔奇和格拉肖的大统一理论基于SU（5）群，这个群包括了SU（3）群——与强核力有关的群；以及SU（2）XU（1）群——与电弱力有关的群。在他们的工作之后，物理学家也曾研究过其他一些可能的大同一群，比如SO（10）以及$E_6$群。

第10章

1.我们将会看到，大爆炸理论中的所谓爆炸，并不是发生在已经存在的空间中的某一位置，所以我们并不会问爆炸究竟发生在哪里这种问题。我们用到的对大爆炸缺点的有趣描述来自艾伦·古斯。可参见他的The Inflationary Uniuerse（Reading, Eng.：Perseus Books，1997），p.Xiii。

2.“大爆炸”这一术语有时被用来代表发生在时间零点——宇宙诞生之时——的事件。但是根据我们在下面的章节中的讨论，广义相对论会在时空零点破产，没有人能够弄清楚时空零点发生了什么。正是因为有这种缺失，我们才说大爆炸理论并不考虑大爆炸的事情。在本章中，我们所有的讨论都局限在方程不会破产的情况下。暴涨宇宙学就是利用这种有良好定义的方程告诉我们，我们想当然的理解成爆炸的那一段空间急速膨胀的过程究竟是怎么回事，而这是大爆炸理论所不能告诉我们的。当然，这种方法还是不能回答宇宙创生之初——如果真有这样的时刻的话——发生了什么。

3.亚伯拉罕·派斯，《上帝是微妙的》（Oxford：Oxford University Press，1982）p.253。

4.对于数学比较好的读者，爱因斯坦将原始方程$G_{\mu\nu}=8\pi T_{\mu\nu}$替换为$G_{\mu\nu}+\Lambda g_{\mu\nu}=8\pi T_{\mu\nu}$，其中Λ代表宇宙常数的大小。

5.在本文中，当我提到物体质量的时候，我指的都是其所有组成粒子合起来的总质量。如果一个立方体，比方说，由一千个金原子组成，那么这个立方体的质量就是单个金原子质量的一千倍。这一定义与牛顿体系相一致。根据牛顿定律，这样的立方体的质量就是单个金原子质量的一千倍，也就是说该立方体比单个金原子重一千倍。而根据爱因斯坦的理论则不是这样，立方体的重量还会依赖于原子的动能（以及其他对立方体的总能量有贡献的量）。这样的结论根据的是公式$E=mc^2$：更多的能量（E），无论其来源为何，都将意味着更多的质量（m）。因而，换一种方式表达这一点就是：因为牛顿不知道$E=mc^2$，所以导致其引力定律中用到的质量定义丢掉了各种能量的贡献，比方说与运动有关的能量的贡献就被丢掉了。

6.这里的讨论对理解深层次物理有启发性，但未能抓住全部要点。压缩的弹簧释放出来的压强的确会对盒子受到的地球引力有影响。但是，这是由于压缩的弹簧影响了盒子的总能量，如前所述，根据广义相对论，总能量才是重要的。不过我在这里解释的要点在于，压强本身——而不只是通过其对总能量的贡献——也会产生引力，就像质量和能量会产生引力一样。在广义相对论的框架下，压强受引力作用。另外，还需要注意的是，我们提到的排斥性引力为充满着具有负压而不是正压的东西的空间区域的内部引力场。在这种情况下，负压贡献的排斥性引力场会作用于区域本身。

7.数学上，宇宙常数用一个数，通常是Λ来代表（参见注释4）。爱因斯坦发现，不管Λ取正还是取负，他的方程都会有完美的意义。正文中的讨论关注的是对现代宇宙学（以及我们将会讨论到的现代天文学观测）有特别意义的情形，也就是令Λ为正数，因为这种取法将带来负压以及排斥性的引力。令Λ为负数将带来通常的吸引性的万有引力。还需要注意的是，由于宇宙常数带来的负压是均匀的，因而这种压强不会直接导致力的产生：只有压强差才能产生力；就好比在水下的时候你的耳膜两边有压强差，所以耳膜会感到压力。而宇宙常数

释放出来的力纯粹就是引力。

8.人们熟悉的磁体总是有北极和南极。而与之相反的是，大统一理论告诉我们，可能存在一个纯为北极或南极的粒子。这样的粒子称为磁单极子。它们会对标准大爆炸宇宙学有很深刻的影响。但人们从未观测到这些粒子。

9.古斯与泰认识到，超冷希格斯场会起到宇宙常数的作用；在此之前，马丁内斯·维特曼和其他人也曾认识到这一点。事实上，泰曾告诉我，要不是《物理评论通讯》（Physical Reuiew Letter）对文章页数有限制，他们就不会在文章末尾加上一句，以说明他们的模型会导致一个指数膨胀时期的出现。但是泰也指出，是古斯首先认识到这样的一段指数膨胀时期（在本章和下一章中将有所讨论）有重要的宇宙学意义，从而一举奠定了暴涨在宇宙学中的核心地位。

有时科学发现的历史会很复杂，俄罗斯物理学家阿列谢·斯塔罗宾斯基在古斯的研究出来之前的几年，通过另一种方法生成了我们所谓的暴涨膨胀，可讲述其研究的论文并未在西方世界赢得广泛的知名度。不过，斯塔罗宾斯基在他的文章中也并没强调这样一段急速膨胀会解决关键的宇宙学问题（比如稍后将会讨论到的视界疑难和平坦性疑难），这就部分的解释了为什么他的研究没能像古斯的工作那样引起强烈反响。1981年，日本科学家也发展了一种暴涨宇宙学，甚至在更早的时候（1978年），俄罗斯科学家甘纳迪·切比索夫与安德烈·林德也有过暴涨的思想，但他们都认识到——仔细研究时就会发现——这样的想法中有一个关键问题（参见注释11）不能克服，因而没有发表他们的研究成果。

数学比较好的读者很容易看出加速膨胀是如何产生的。爱因斯坦方程之一为$d^2a/dt^2/a=-4\pi/3(\rho+3P)$，其中a，ρ，P分别为宇宙的标量因子（其“大小”），能量密度以及压强。注意如果方程的右边为正，标量因子就会加速增加：宇宙的增长率就会随时间变大。对于盘

踞在势能碗中央高地的希格斯场来说，其压强密度就会变得等于负的能量密度（宇宙常数也是这样），这样一来方程的右边就只能是正的了。

10.这些量子跃迁背后的物理是第4章讲过的不确定原理。我会在第11章以及第12章中清楚的讨论量子不确定性在场上的应用，但我先在这里简要的讲解一下作为预习。空间中给定点处的场值，以及给定点处的场值变化率，对于场来说至关重要，就如位置与速度（动量）对于粒子至关重要一般。因而，正如我们不能同时知道一个粒子的位置与速度；在空间中的任何一点，一个场也不会同时有确定的值及确定的场值变化率。某一时刻的场的值越为明确，其场值的变化率就越不确定——换句话说，下一时刻场值发生改变的可能性就越大。而量子不确定性带来的这种变化，就是我要说的场值中的量子跃迁。

11.林德、阿尔布莱奇与斯坦哈特的贡献绝对是非常重要的，因为古斯的原始模型——现在称为旧暴涨理论——有一个严重的缺陷。还记得吗？超冷希格斯场（或者，我们按照专业术语将之称为暴涨子场）的一个值位于其能量碗的中央高地，这个值在整个空间中具有均一性。这样一来，因为之前我已经讲过超冷希格斯场可以极快地跃迁到其最低能量值，所以我们就得问问，这样的由量子效应导致的跃迁有没有可能于同一时刻在整个空间发生？答案是否定的。相反的，正如古斯提出的那样，希格斯场释放到零能量值的过程会通过所谓的泡沫核化（Bubble Nucleation）过程发生：暴涨子场在空间中某点处的场值变为零，产生了一个向外延展的泡沫，这个泡沫的外壳以光速运动，泡沫所过之处，暴涨子场的值都跌落为零。在古斯眼中，正是这随机出现在任意位置的大量泡沫最终使宇宙空间中的所有地方的暴涨子场的场值都跌落为零。这里的问题，古斯本人也意识到了，在于围绕着泡沫的空间中仍有非零能量的暴涨子场，这样的区域仍会急速暴涨膨胀，驱使泡沫彼此远离。因而，根本没法保证所有正在变大的泡沫会合在一起形成巨大的均匀的空间膨胀。而且，古斯提出，在暴涨

子场的能量跌落为零的过程中，其能量并未丢失，而是转化为普通的物质及辐射粒子留驻宇宙。为了使模型能同实验观测相符合，这种转化必须带来均匀的遍布于整个空间的物质与辐射分布。在古斯提出的机制中，这种转化会通过泡沫外壳的碰撞而得以发生，但是计算——古斯与哥伦比亚大学的埃里克·温伯格做了这个计算；剑桥大学的史蒂芬·霍金，依安·摩斯以及约翰·斯图尔特也做了该计算——表明，如此得到的分布并不均匀。而且，细细推敲古斯的原始暴涨模型会发现很多重大问题。

林德、阿尔布莱奇与斯坦哈特的才智——现在成为新暴涨理论——解决了这些令人头疼的问题。这几位科学家将势能碗的形状变成了图10.2中的样子，这样一来，暴涨子场就会自然地从能量山上“滚”落到能量谷中，得到零能量值，这样渐变适度的过程并不需要原始理论中的量子跃迁。而且，正如这几位科学家通过计算证明的那样，这种渐变的滚落至谷底的过程充分延长了空间的暴涨膨胀，以至于单个的泡沫能够轻而易举的变大到包容整个可观测宇宙的程度。因而，在这种方法中，根本没必要担心泡沫的结合问题。具有同等重要性的是，在旧暴涨理论中，暴涨子场中的能量通过泡沫的碰撞转化为普通的粒子与辐射，而在新暴涨理论中，暴涨子逐渐的在全空间中统一的完成这个能量转化过程，整个过程类似于摩擦：随着从能量山上滚落——在整个空间中统一的滚落——暴涨子场通过与熟悉的粒子与辐射的场之间的“磨蹭”（相互作用）将能量传递给粒子与辐射。因而，新暴涨理论保留了古斯理论的所有成功之处，并同时修补了旧理论的重大问题。

在新暴涨理论带来重要进展差不多一年之后，安德烈·林德取得了另一个突破。新暴涨要想成功实现，一系列关键元素必须得以满足：势能碗必须有正确的形状，暴涨子场的值一开始必须在势能碗的高位（更技术化点说，暴涨子场值本身在足够大的空间膨胀中必须保持均匀性）。尽管宇宙有可能满足这样的条件，林德却找出了一种更为简

单，也更少人为设计的方式实现暴涨膨胀。林德认识到，即使用一个较为简单的势能碗，比如图9.1（a）中所示的那样；即使不精细调节暴涨子场的初始值，暴涨也有自然发生的可能。他的想法是这样的。想象一下，在极早期宇宙中，一切都处于“混沌”状态——比如说，很可能有一个暴涨子场，其场值随机的变来变去。在空间中的某些位置，其场值可能很小，而在另一些位置，其场值可能中等，还有一些位置，其场值可能很高。于是，在场值很小或者中等的位置处可能就没什么特别值得注意的事情发生。林德认识到，在场值很高的位置处（即便该区域很小，比如只有$10^{-33}$厘米见方也没关系），一些极其有趣的事情可能会发生。当暴涨子场的场值很高的时候——当暴涨子场位于图9.1（a）中的势能碗的高位时——一种宇宙摩擦开始起作用了：场值倾向于滚下山到较低势能处，而其高场值又会导致起阻碍作用的力产生，在这种力的作用下场值下滑速度很慢。这样一来，暴涨子场的值就几乎保持不变（就像新暴涨理论中的势能山顶部的暴涨子一样），并会贡献一个几乎恒定的能量和一个几乎恒定的负压强。现在我们已经很熟悉了，暴涨膨胀必须在一些条件得以满足的情况下才会发生。因而，既不需要有一个特别形状的势能碗，也不需要特别设置暴涨子场的位型，早期宇宙的混沌环境就可以自然的诱发暴涨膨胀。所以不用奇怪，林德将这种方法称为混沌暴涨。很多物理学家将此视为最为可信的暴涨理论。

12.对这段历史比较熟悉的读者应该认识到，古斯的发现之所以令人振奋，其原因在于——正如我们简要提过的那样——它能解决重要的宇宙学问题，比如视界疑难以及平坦性疑难。

13.你可能会想知道，电弱希格斯场，或是大统一希格斯场会不会扮演双重角色呢？既扮演我们在第9章中讲过的角色，又要在希格斯海形成之前的早期宇宙中，驱动暴涨膨胀。研究人员的确提出过一些此类模型，但这些模型通常都有一些克服不了的技术问题。暴涨膨胀最令人信服的实现方式就是用一个新的希格斯场扮演暴涨子场的角

色。

14.参见本章注释11。

15.比方说，你可以将我们的视界想象为巨大的球面，而我们位于球心；自从大爆炸以来，这个球面就将那些我们能够与之联系的区域（球面内的区域）与我们无法与之联系的区域（球面外的区域）分割开来。今天，我们的“视界球”的半径差不多是140亿光年；在整个宇宙历史的前期，这个半径要小得多，因为可供光传播的时间并不多。亦可参见第8章注释10。

16.这正是暴涨宇宙学解决视界疑难的关键所在，为避免混淆，让我来就解决办法的关键步骤做一番提示。如果某天晚上，你和你的好朋友在旷野中通过开关灯来交换光信号，你们玩得很高兴。你会发现，不管你们切换信号以及跑离彼此的速度有多快，你们总是可以交换信号下去。为什么会这样？这个嘛，要是想收不到你朋友照向你的光，或者你朋友想收不到你照过去的光，你们彼此远离的速度就要大于光速，而这是不可能的。那么，为什么在宇宙初期（因而具有相同的，比如说，温度）还可以交换光信号的空间区域，到了今天就会发现彼此超出了对方的沟通范围呢？你和你朋友的例子就能说明这个问题，答案就是，再也无法联络彼此的区域一定是以大于光速的速度远离彼此。而事实上，暴涨阶段的排斥性万有引力那巨大的外推力的确能够驱动不同的空间区域以大于光速的速度远离彼此。再说一遍，这与狭义相对论并不矛盾，因为光所设定的速度极限是相对于穿越空间的运动而言的，而不是相对于空间自身的膨胀而言的。所以，暴涨宇宙学一个新颖且重要的特性即在于，在一个极短时期内，空间超光速膨胀。

17.注意，临界密度的数值会随着宇宙膨胀而减小。但问题的关键在于，如果在某个时刻宇宙实际的质量或能量密度等于临界密度，那么实际密度就会按照与临界密度完全一样的方式减小，并始终与临

界密度保持相等。

18.数学比较好的读者应该会注意到，在暴涨阶段，我们宇宙视界的大小固定不变，而空间膨胀了很多（将第8章注释10中的标度因子取为指数形式，就能轻易地看出这一点）。正是因为这样，所以在暴涨理论的框架下，我们可观测的宇宙只是巨大宇宙中的一个小斑点。

19.R.普莱斯顿，First Light（New York：Random House Trade Paperbacks，1996），p.118。

20.若想一般性的了解一下暗物质，可以参看L.克劳斯，Quintessence：The Mys-tery of Missing Mass in the Uniuerse（New York：Basic Books，2000）。

21.专业读者应该会看出，我并没有区分各种来自于不同尺度（星系尺度、宇宙尺度）的观测的暗物质问题，因为暗物质对宇宙质量、密度的贡献是我在这里唯一的兴趣点。

22.实际上，对于这是否就是所有Ia型超新星背后的物理机制（我在这里感谢为我指出了这一点的D.思博格尔）这一点，仍存在着某些争议，但是这些事件的一致性——这才是我们需要讨论的——却有扎实的观测证据。

23.非常有趣的一点是，早在得到超新星的结果的很多年前，在普林斯顿大学的吉姆·皮伯尔斯、凯斯西储大学的劳伦斯·克劳斯，以及芝加哥大学的迈克尔·特纳等人所做的预见性的理论工作中，就提出宇宙可能会有一个小的非零宇宙常数，那时，很多物理学家并没有认真对待这种观点，但到了今天，有了超新星数据后，人们的态度发生了巨大的变化。同样需要注意的是，我们在本章的前面部分看到，宇宙常数的外推力可由希格斯场模拟；因为希格斯场，就像碗中高处的

青蛙那样，可以盘踞在其最低能量之上的位置。所以，虽然宇宙常数可以符合实验数据，但超新星研究人员所得出的更为准确的说法却应该是：空间中必然充斥着如宇宙常数一般能生成外推力的东西（希格斯场也可以产生长期的外推力，而不是只能产生暴涨宇宙学中原初时刻短暂的外推力。我们将在第14章中讨论这一内容，在那里，我们将要考虑一些类似于数据是否真的需要宇宙常数来解释，或者是否存在具有相同引力效应的其他实体这样的问题）。研究人员常常用“暗能量”这一术语来指代一种宇宙组分，这种组分不可见，但却会使空间中的区域彼此之间产生推力，而不是拉力。

24.用暗能量来解释观测到的加速膨胀为人们普遍接受，但是也有一些理论走得更远。比如说，在某些理论中，在极大的尺度上——宇宙学尺度上，引力的大小与牛顿理论以及爱因斯坦的理论所预言的强度有一定偏差，而这种偏差就可以用来解释所观测到的实验数据。还有一些人并不相信实验数据意味着宇宙加速膨胀，他们认为需要有更加精确的数据来确认这一点。在脑海中记下这些其他的想法很重要，特别是未来的观测可能会改变现有的解释这一点。但是目前来说，大多数科学家相信正文中所描述的理论解释。

第11章

1.20世纪80年代早期，指出量子涨落是如何带来空间上的各向异性的科学家以下列人物为主：史蒂芬·霍金，阿列谢·斯塔罗宾斯基，艾伦·古斯，So-Young Pi（韩裔，英文音译），詹姆斯·巴登，保罗·斯坦哈特，迈克尔·特纳，维亚切斯拉夫·马克哈诺夫以及甘纳迪·切比索夫。

2.即使有了正文中的讨论，你可能还是会对一个暴涨子小块中的那么少量的质量或能量究竟怎样才能导致可观测宇宙中那么大量的质量或能量感到迷惑。究竟怎样才能在结束的时候搞出那么多质量或能量呢？这个嘛，正如在正文中解释过的，暴涨子场，利用其负压，从

引力中“挖掘”出了能量。这就意味着，随着暴涨子场中的能量增加，引力场中的能量将会减少。早在牛顿时代即已为人所知的一个引力场性质是，其能量可以达到任意大小的负。因而，引力场就像一家愿意无限贷款的银行——引力能够提供无限的能量，这些能量就是暴涨子场在空间膨胀时抽取出来的。

均匀的暴涨子场的初始核特定的质量和大小取决于人们所研究的暴涨宇宙学模型的细节（尤其是，暴涨子场的势能碗的准确形状）。在正文中，我假定初始暴涨子场的能量密度为每立方厘米$10^{82}$克，这样的话，（$10^{-26}$厘米）3＝$10^{-78}$立方厘米内的总质量就有10千克，也就是20磅之多。这是传统型暴涨子模型中的典型数值，但只是为了令你对所涉及的数字有一个粗浅的直观感觉。为了了解一些可能的范围到底有多大，我们来看看安德烈·林德混沌暴涨模型（参见第10章注释11），其中，我们的可观测宇宙来自于一个还要小的初始核，横竖只有$10^{-33}$厘米左右（所谓的普朗克长度），而其能量密度则更高，达到每立方厘米$10^{94}$克，将这样的数值组合起来我们可以看出，总能量将只有$10^{-5}$克（所谓的普朗克质量）。在这些暴涨理论中，初始核就像一粒尘埃那么重。

3.参见Paul Davies的文章“Inflation and Time Asymmetry in the Universe”，刊登于Nature，301卷398页；Don Page的文章“Inflation Does Not Explain Time Asymmetry”，刊登于Nature，304卷39页；以及Paul Davies的文章“Inflation in the Universe and Time Asymmetry”，刊登于Nature，312卷524页。

4.为了解释清楚本质所在，我们最好将熵想成两部分，一部分来自于时空和引力，另一部分可归结为其他的一切，这样我们就能在直观上抓住要点。但是，我得指出的是，给出一个数学上严格的处理——其中，引力对熵的贡献被明确地找到，分离出来，并加以解释——是非常困难的。不过，这并不会危及我们的定性结论。一旦遇到这样的麻烦，你要知道的是，整个讨论的很大部分都可以在不涉及引

力熵的情况下重新表述。如我们在第6章中强调过的那样，当有关的是普通的吸引性的引力时，物质聚团。这样的话，物质将引力势能转化为其自身的动能，接着，其中的一部分动能又以辐射的形式从物质团中释放出来。这是一个熵增的事件（粒子平均速度越大，相应的相空间体积就越大；相互作用导致的辐射产生增加了总的粒子数——而辐射和粒子都会使总的熵增加）。这样一来，我们在正文中所说的引力熵就可以重新表述为引力导致的物质熵。当我们说引力熵很低的时候，我们指的是引力使物质聚团而产生的大量熵。在实现这样的熵的过程中，物质的聚团导致了不均匀的、非各向同性的引力场——时空的蜷曲和褶皱——而在正文中，我将其描述为有了较高的熵。但是，随着讨论变得清楚，我们实际可以这么想，物质越是聚团（以及过程中产生的辐射），就越会有较高的熵（相比于均匀分布时）。这点非常的好，专家级读者可能会注意到，如果我们将经典引力背景（经典时空）视为引力子的相干态，那它本质上就是一种独一无二的态，因而具有低熵。只有将问题适当粗化，我们才有可能得到熵的排布。正如这条注释所强调的，这并没有什么特别的必要性。另一方面，要是物质聚团足以产生黑洞，那就会有一种无可辩驳的熵的排布方式：黑洞视界的表面积将正比于黑洞的熵（我们将在第16章中探讨这些问题）。而这种熵，毫无疑问，应该被称为引力熵。

5.正如鸡蛋破碎和破碎的鸡蛋再变回完好的鸡蛋这两种可能性都存在；量子导致的涨落长成较大的各向异性（如我们讲过的）或充分关联的各向异性协力压低这种增长也都具有可能性。因而，要想用暴涨解释时间之箭，就得要求初始的量子涨落之间没有足够的关联度。再说一遍，如果我们按玻尔兹曼式的方式思考，那么在所有能导致暴涨发生条件的涨落中，迟早有一个会真的满足这个条件，导致我们知道的这个宇宙开动起来。

6.有一些物理学家会宣称情况比我们讲的要好一点。比如说，安德烈·林德就认为，在混沌暴涨中（参见第10章注释11），可观测宇

宙来自于普朗克尺度大小的硬核，这个硬核中包含着具有普朗克尺度能量密度的均匀暴涨子场。在这种假设下，林德进一步提出，在如此之小的硬核中的均匀暴涨子场的熵差不多等于任意其他暴涨子场的熵，因而，暴涨所需的条件实在没什么特别之处。普朗克尺度的硬核中的熵如此之小，但是却与普朗克尺度的硬核中可能具有的熵一样大小。期待中的暴涨膨胀一下子产生了，一瞬间，创造出了一个有着极高熵的巨大宇宙，但是由于其平滑均匀的物质分布，这个宇宙中的熵仍远远未到它本可能有的熵的量。时间之箭所指的方向正是可填平这巨大的熵壑的方向。

虽然我对这乐观的看法有些偏心，但不能忘了小心谨慎，除非我们对诱发暴涨物理有了更好的把握。比如说，专家级读者应该会注意到，用到了那些有关高能（普朗克级）场模式——可以对暴涨的开始有所影响并会在结构形成中扮演重要角色的模式——的很好但却未经证明的假设。

第12章

1.我在这里能想到的间接证据与这样一个事实有关：除引力之外的3种力的强度取决于力起作用的环境的能量和温度。在能量和温度很低的时候，比如在我们日常生活的环境中，3种力的强度各不相同。但是有间接的理论和实验证据表明：当温度很高的时候，比如宇宙的最初时刻，3种力的强度将趋近于一点；这就相当于间接告诉我们，这3种力在基本层面上很有可能是统一的，只是到了能量和温度很低的时候才显示出区别。更为详尽的讨论可参见《宇宙的琴弦》第7章。

2.我们知道某种场——比如任何一种已知的力场——是宇宙组成中的一种，我们就知道这种场无处不在——宇宙中的每个角落都有它的身影。我们没法将场驱除掉，就像我们没法驱除空间本身一样。我们最多只能使它们取其能量最小时的场值。对于力场——比如电磁场

——而言，这个值就是零——我们在正文中有所讨论。对暴涨子场或标准模型中的希格斯场（为简要起见，我们不在这里讨论这种场）来说，这个值就可能不是零，具体是多少要取决于其势能的准确形状——参见我们在第9章、第10章中的有关讨论。如我们在正文中所说，为了使我们的讨论顺畅，姑且只讨论那些一旦场值为零即达到最低能量的场的量子涨落。我们所得到的结论也可以不经修改直接推广到希格斯场和暴涨子的涨落。

3.实际上，数学不错的读者应当注意到，不确定原理表明：能量的涨落反比于我们测量的时间分辨率，所以，我们测量场的能量所用的时间分辨率越精细，场的涨落就越大。

4.在这个验证卡西米尔力的实验中，拉莫雷奥克斯修改了实验设置：考察的对象变成了球面透镜与石英片之间的吸引力。更加晚近的时候，意大利帕多瓦大学的詹尼·卡鲁格诺、罗伯托·奥诺佛里奥及其合作者采用了更为困难的原始卡西米尔实验设置——两个平行板——再次做了这个实验（使两块板保持完美的平行状态是一项艰巨的实验挑战）。到目前为止，这组人在15%的水平上证实了卡西米尔的预言。

5.回头来看，即使爱因斯坦没在1917年引入宇宙常数，量子物理学家们也会在几十年后按他们自己的方式引入宇宙常数。你可能还记得，爱因斯坦——以及当代的宇宙常数支持者们——将宇宙常数视为一种弥漫于整个空间的能量，但却无法搞清楚这种能量的起源。我们现在知道，量子物理在空间中塞满了起伏不定的场；而且，我们还可以通过卡西米尔的发现直接看到，这些微观场在空间中塞满了能量。事实上，理论物理面对的一大挑战就是要证明所有场的涨落的联合贡献在真空中产生的总能量——总宇宙常数——并没有大于超出在第10章讨论过的超新星观测所带来的限制。到目前为止，还没有人成功地做到这一点。精确的理论分析又早被证明超出了目前的理论能力；而近似计算得到的结果比观测大很多，又表明近似错得厉害。很多人都

将解释宇宙常数的值（是否如长久以来认为的那样为零，或是暴涨及超新星数据暗示的很小但却非零）视为理论物理尚未解决的重大问题之一。

6.我将在本节中讲一种能看出量子力学与广义相对论的矛盾的方法。但是为了紧跟我们寻找空间和时间真正性质这一主题，在尝试调和广义相对论与量子力学的过程中，我也会把注意力放到一些不那么切实的但却具有潜在重要性的谜题上。将经典的非引力理论（比如麦克斯韦的电动力学）转换为量子理论的程序，如果被直接用来转换经典的广义相对论（比如布莱斯·德维特的工作——现在被称为惠勒德维特方程），那么就会有一些极有吸引力的问题出现。在这类理论的核心方程中，时间变量并不会显示出现。所以，在这样量子化引力的方法中，并不需要数学上明显的出现时间——就像其他的基本理论所必需的那样。时间上的演化通过我们认为在常规方式下会有所改变的宇宙物理性质（比如说其密度）得以体现。至今还没人知道这种量子化引力的方式恰当与否（尽管在这个体系的一个支流上，圈量子引力，已经取得了一些进展；参见第16章），所以人们并不清楚明显时间变量的缺失是否是某种深层次的暗示（时间是衍生概念吗）。在本章中，我们的注意力将集中于另一种调和广义相对论和量子力学的方法——超弦理论。

7.黑洞“中心”这种说法有点用词不当，就好像空间中真有这么个位置似的。而大体上讲，原因在于，当越过黑洞视界——黑洞外边缘——时，空间和时间的角色互换了。事实上，正如你不能停在一秒，不进入下一秒；你一旦越过黑洞视界，就会被拖进黑洞“中心”。在时间上向前进和在黑洞中向着中心去的这种类比关系是受黑洞的数学描述所启发。因而，不要把黑洞中心想成空间中的一个位置，最好要把它想成是时间上的一个位置。而且，因为你逃不脱黑洞的中心，所以你可能会认为黑洞的中心在时空中的位置就是时间的尽头。这可能是对的。但是，因为标准的广义相对论方程在这样极端的质量密度下不

再成立，我们不太有做这种明确陈述的能力。很明显，这意味着我们有了不会在黑洞中破产的方程，我们有可能洞察时间的重要性质。而这是超弦理论的目标。

8.同前几章一样，当我说“可观测宇宙”时，我指的是，在大爆炸以后，我们可以与之联系——哪怕只有理论上的可能性——的宇宙部分。在一个空间上无限大的宇宙中，比如第8章中所讨论的宇宙，在大爆炸的那一刻，所有的空间并不会缩成一个点。当然，我们越是往回看，宇宙可观测部分的一切就会被挤压到越小的空间中，但是，虽然很难刻画，可就是有一些事物——距离我们无限远的事物——与我们永远的分割开来，不管物质和能量密度变得多高都将是这样。

9.莱昂纳德·萨斯金，见“优雅的宇宙”，NOVA，3小时PBS系列节目，2003年10月28日与11月4日首播。

10.事实上，设计检验超弦理论的实验时所遇到的苦难，一直是导致超弦理论不被接受的重要障碍。但是，如我们在后面的章节中将会看到的那样，在这个方向上已经取得了很大进展；即将到来的加速器实验以及太空实验很有希望为超弦理论提供至少是间接的支持证据，要是够幸运的话，可能还不止如此。

11.尽管我没在正文中清楚地讲到，但有必要知道，所有的已知粒子都有反粒子——具有相同质量相反力荷（比如相反符号的电荷）的粒子。电子的反粒子就是正电子，上夸克的反粒子当然就是反上夸克，以此类推。

12.我们在第13章中将会看到，弦论中的一些最新工作表明，弦可能会比普朗克长度大很多，而这会带来很多重要的影响——比如使理论变得可通过实验验证。

13.原子的存在最初就是通过一些间接方法讨论（比如对各种化

学物质按一定比率组成的解释，以及之后对布朗运动的解释）；黑洞的存在最初也只能通过间接效应——其附近星体的气体落入其中——得以确认，而不能直接“看到”。

14.既然一根轻微振动的弦也有一定质量，你或许会想知道，弦的振动模式是否有可能导致零质量粒子。答案——再次强调一下——与量子不确定性有关。不管弦有多么安静，量子不确定性都会使其产生最小水准的涨落。而且，量子力学的古怪之处在于，这些不确定性导致的涨落具有负的能量。将这种负能量与来自于普通弦的最轻微的振动导致的正能量合在一起，总质量或总能量就是零。

15.对数学比较好的读者来说，较为准确的说法应该是，弦的振动模式的质量平方等于普朗克质量平方的整数倍。更加准确的说法是（与第13章中要讲的近期进展有关），这些质量的平方等于弦的标度（反比于弦的长度的平方）的整数倍。在传统弦论体系中，弦的标度与普朗克质量非常接近，这就是在正文中我只简单地说普朗克质量的原因。但是，在第13章中，我们将讨论弦的标度与普朗克质量不同的情况。

16.即使只用简单的术语也不难理解，普朗克长度是怎样进入克莱因的分析中的。广义相对论和量子力学一共使用了3个常数：c（光速），G（引力的基本强度）以及h（刻画量子效应大小的普朗克常数）。这3个数组合到一起可以产生一个有长度单位的量：$(hG/c^3)^{1/2}$，根据定义，这就是普朗克长度，将这3个常数的数值代入后，我们发现普朗克长度大约是$1.616\times10^{-33}$厘米。因而，除非理论中能够出现一个与1差别很大的无量纲数——而这样的数在一个简单的、体系良好的物理理论中并不常见——否则我们就会认为普朗克长度就是长度（比方说蜷曲维度的长度）的特征大小。不过，要注意的是，这并不会排除掉维度比普朗克长度大很多的可能性，我们在第13章中将会看到严格探讨这种可能性的有趣进展。

17.使一个粒子具有电子电荷，但同时却有相对较小的质量，早被证明为不可能的任务。

18.注意，促使我们在第8章中用对称性要求来限制宇宙形状的是3个大维度中的天文学观测（比如微波背景辐射）。这些对称性对可能存在的6个额外维度没有影响。图12.9（a）基于安德鲁·汉森的构想。

19.你可能会想知道，除了空间有额外维度外，时间能不能有额外维度。研究人员（比如南加利福尼亚大学的伊特扎克·巴斯）探索过这种可能性，研究结果证明构建一个具有第二个时间维度的理论看起来在物理上是合理的。但这第二个时间维度到底与本来的时间维度同等重要还是只是某种永无实际意义的数学产物呢？一般的看法趋向于后者。而与之相反的是，最直接的弦论文章都会告诉我们，额外维的空间维度一丝一毫都像我们平常感受的三维空间那样真实。

20.弦论专家（以及那些读过《宇宙的琴弦》第12章的读者）将会认识到，更为准确的说法应该是，某些弦论体系（将在本书的第13章中讨论）允许与十一维有关的极限。弦论最好应该被想成是基本层面上的十一维时空理论呢？还是应将十一维时空体系视为某种与其他极限具有同等地位的某种特定极限（比如说，在IIA型理论中，将弦的耦合常数取得很大）呢？在这点上仍有很多争论。这个方向上的讨论与我们在一般水平上的讨论关系不大，而我之所以采用第一种观点，在很大程度上是因为确定又一致的总维数在语言上比较简便些。

第13章

1.对于数学比较好的读者来说，我在这里讨论的是共形对称性——这一对称性指的是对时空中的体积所做的任意保角变换可有假设的基本组分消除。弦要占用两维时空面，弦论的方程在两维共形群——一个无限维度的对称群——下具有不变性。与之相反，对于其他

数目的空间维度——与本身不是一维的物体相关联——共形群都只是有限维。

2.对这些发展作出重大贡献（要么做了奠基性工作，要么作出了后续发现）的科学家有很多：麦克尔·达弗，保罗·霍维，稻见武夫，凯利·斯黛拉，埃里克·博格舒夫，埃尔金·赛金，保罗·唐森，克里斯·赫尔，克里斯·蒲柏，约翰·施瓦茨，艾索科·森，安德鲁·斯特劳明格，柯蒂斯·卡兰，乔·波金斯基，皮特·哈罗瓦，戴瑾，罗伯特·利，赫尔曼·尼克莱，伯纳德·德维特，以及另外一些没有提到名字的科学家。

3.事实上，如在《宇宙的琴弦》第12章中解释过的那样，第10个空间维和P-膜之间还有一种更为紧密的联系。当你增加某个理论，比如说IIA型弦论中的第10个空间维度的大小时，一维的弦延展成两维的类似于管子内部形状的膜。如果你假定第十维很小，就像总与这些发现无关似的，内管看起来就像弦似的，并且其行为也像弦似的。同弦的情况一样，这些膜是否不可再分，或者换句话说，是否由更加精细的成分组成，仍然没有答案。根据现有的认识，弦论或M理论中的成分已经是宇宙的最基本组分，然而，还有更加基本组分存在的可能性并未完全堵死。因为接踵而来的很多东西已经与我们的问题相距太远，所以在我们的讨论中，将只采用最简单的观点，即现有的这些组分——弦以及各种维度的膜——已经是最基本成分。那么，前面的讨论所得出的结论——更高维度的基本客体没法被纳入物理上合理的框架之中——又是怎么回事呢？这个嘛，先前的推理本身植根于另一种量子力学近似方案，这种方案虽然标准化且有充分的实验依据，但就像任何近似一样，都是有其局限性的。尽管研究人员尚未弄清楚将更高维度的客体纳入量子理论有关的所有细节，但是，这些高维成分却与5种弦论都那么适合又那么自洽，以至于几乎每个人都相信不可能存在对基本又神圣的物理定律的可怕破坏。

4.事实上，我们有可能生活在更高维度的膜上（4-膜，5-膜……），其中3个维度为普通空间，而其他的维度是理论所要求的

更小的额外维度。

5.数学比较好的读者应该会注意到，弦论学家早就知道闭弦遵从所谓的T对偶性（在本书的第16章，《宇宙的琴弦》第10章中有所解释）。基本上，T对偶性说的是，如果所谓额外维度的形状都是圆环，那么弦论将完全无法知道圆环半径是R还是1/R。原因在于，弦既可以绕着环运动（“动量模式”），也可以盘绕在环上“缠绕模式”，将R与1/R对换，物理学家们认识到只是这两种模式所扮演的角色互换了，而理论的整体物理性质并没有发生变化。对这串推理重要的是弦得是闭弦；因为对于开弦来说，没有拓扑稳定的缠绕于圆环维度的概念。所以，乍看之下，开弦和闭弦在T对偶下表现得完全不同。而利用开弦的狄利克雷（D膜中的D）边界条件进一步分析后，波金斯基、戴瑾、利以及哈罗瓦、格林和其他一些研究者解开了这一谜题。

6.一些试图避免引入暗物质与暗能量的方案提出，在大尺度上被广为接受的引力行为也可能与牛顿和爱因斯坦认为的有所不同，他们就是这样解释所看到的物质与引力效应的不符。但是，这不过是些猜想性的方案，既没有实验根据也没有理论支持。

7.引入这个想法的物理学家是S.吉丁斯、S.托马斯、萨瓦斯·蒂莫普洛斯以及G.兰斯伯格。

8.注意，这样脉动宇宙的收缩阶段与反过来的膨胀阶段是不一样的。物理过程，比如鸡蛋破碎、蜡烛融化等，在膨胀阶段中普通的“向前”时间方向上会发生，在接下来的收缩阶段会继续发生。这就是为什么熵在两个阶段中都是增加的。

9.专家读者应该会注意到，循环模型可以用某张3-膜上的四维有效场论的语言表述，在这种形式下，循环模型会有一些更为人所熟悉的标量场驱动的暴涨模型的性质。当我说“激进的新机制”时，我想说的是，用碰撞的膜的术语的概念性描述，而碰撞的膜模型本身就是一

种全新的思考宇宙学的方法。

10.不要数维数数糊涂了。两张3-膜，以及其间的空间，共有4个维度，加上时间就是5个。这样就为卡拉比丘流形留有6个维度。

11.一个重要的例外，我们在本章的结尾提到并将在第14章进一步加以讨论，与引力场中的各向异性有关，是所谓的原初引力波。在这一点上，暴涨宇宙学与循环宇宙模型有所不同，因而这是一种可以通过实验区分两种理论的办法。

12.量子力学保证了总会有一个非零概率使得偶然的涨落摧毁循环过程，从而导致模型慢慢停下来。即便这一概率很小，它也迟早会发生，因而循环不能无限期进行下去。

第14章

1.A.Einstein，“Vierteljahrschrift für gerichtliche Medizin und ?ffentliches Sanit?tswesen”4437（1912）。D.Brill and J.Cohen, Phys.Reu.Vol.143，no.4，1011（1966）；H.Pfister and K.Braun, Class.Quantum Grau.2，909（1985）.

2.在席夫和普夫提出他们的想法之后的40年间，人们也做了很多其他的关于框架曳引的实验。这些实验（由布鲁诺·伯托蒂、伊格纳奇奥·丘弗里尼、彼得·本德领导的实验组，以及I.I.夏皮罗、R.D.里森博格，J.F.钱德勒，R.W.拜伯科克领导的实验组）研究了月球以及绕地球运行的卫星的运动，并且从中发现了一些框架曳引效应的证据。引力探测器B卫星的优势在于，它是这方面第一个真正周详的实验，完全在实验学家的掌控之下，因而将给出框架曳引最精确最直接的实验证据。

3.尽管这些图片可以非常有效的带给读者有关爱因斯坦发现的直观感受，但它们还有其他的局限性：不能展示时间蜷曲。而这一点非

常重要，因为根据广义相对论，像太阳这样的普通物体，与黑洞之类的极端情况不同，时间蜷曲（你越靠近太阳，你的钟就走得越慢）远比空间蜷曲来得明显。在纸上画出时间蜷曲并不容易，在纸面上反映出时间蜷曲如何对地球绕太阳的椭圆轨道之类的弯曲空间轨道产生影响也不简单，这正是图3.10（我所看到过的有关广义相对论的图片大抵如此）只展示空间蜷曲的原因。但我们必须在脑中牢记：在很多普通的天体物理环境中，时间蜷曲更为主要。

4.1974年，拉塞尔·哈斯和约瑟夫·泰勒发现了双脉冲星——绕着彼此运动的两个脉冲星（急速旋转的中子星）。因为这两个脉冲星运动得很快并且靠得很近，所以根据广义相对论的预言，这两个脉冲星将会释放出巨量的引力辐射。尽管很难直接观测到这种辐射，但是广义相对论告诉我们，这种辐射将通过其他办法展示自身的存在：辐射带走的能量将使这两个脉冲星的轨道周期逐渐衰减。自其被发现后，人们就一直在观测这样的双脉冲星，观测结果的确显示出其周期在衰减——并且在千分之一的水平上与广义相对论的预言相符合。因而，即使我们不能直接探测到引力辐射，我们还是有其存在的有力证据。因为双脉冲星的发现，哈斯和泰勒于1993年被授予诺贝尔物理学奖。

5.但是，见上面这条注释4。

6.从动力学的角度看，宇宙线可算是天然的加速器，并且这台加速器的能量远比我们现有的和短期内会拥有的加速器的能量高。它的缺点在于，虽然宇宙线中的粒子能量非常高，但我们却完全无法操控粒子碰撞——一旦涉及宇宙线碰撞，我们只能当个被动的观测者。而且，给定能量的宇宙线粒子的数目会随着能级的增加而减少。要是每秒钟地球表面每平方千米的范围内进入一百亿个能量等价于质子质量（相当于大型强子对撞机设计能量的千分之一）的宇宙线粒子的话，那么每个世纪撞入地球表面每平方千米的范围内的最高能量（相当于千亿倍质子质量的能量）的粒子只有一个。最后，加速器可以让粒子沿相反方向快速碰撞，从而获得很高的质量能量。但是宇宙线粒子，

只能撞上相对能量低得多的大气粒子。不过，这些缺点也并不是不能克服。依靠过去几十年的努力，实验学家们已经通过研究更加丰富的低能宇宙线数据掌握了很多本领，而且，为了对付那极少量的高能碰撞，实验学家们已经建造了大批量的探测器来捕获尽可能多的粒子。

7.专业读者可能会认识到在一个具有动态时空的理论中能量守恒问题非常深奥。当然，爱因斯坦方程所有源的应力张量具有协变不变性。但这并不一定意味着整体的能量守恒律。有理由认为，应力张量并不对应着引力能量——广义相对论中人所共知的艰深概念。在足够短的距离和时间尺度上——比如加速器实验的距离和时间尺度——局域能量守恒肯定有效，但却要小心处理整体能量守恒的问题。

8.这里指的是最简单的暴涨理论中的情况。研究人员已经发现，在暴涨理论更为复杂的版本中，这种引力波的产生可能会被压低。

9.可行的暗物质候选者必须得是稳定或者说长寿的粒子——不能分解成其他粒子。最轻的超对称粒子就有这种性质，因而，更准确的说法是zino、higgsino以及Photino中最轻的一个将是合适的暗物质候选者。

10.不久之前，一个在意大利的Gran Sasso实验室工作的意大利中国联合研究组（暗物质实验组，Dark Matter Experiment, DAMA）宣布了一个令人兴奋的消息：他们首次成功的直接探测到了暗物质。但是，直到目前为止，还没有其他的实验组能够证实该组宣布的消息。事实上，另一个实验，坐落于斯坦福由美国以及俄罗斯的研究人员共同参与的低温暗物质探寻（Cryogenic Dark Matter Search, CDMS）已经采集了大量的数据，很多人相信CDMS的实验数据已经在很高的置信度上排除了DAMA的结果。除了这几个，还有很多其他的寻找暗物质的实验正在进行之中。要想阅读相关的资料，可以看看这个网址httP://hePwww.rl.ac.uk/ukdmc/dark_matter/other_searches.html。

第15章

1.这里的说法忽略了隐变量理论，比如玻姆的理论。但即使在这样的理论中，我们想要传输的也是物体的量子态（波函数），所以仅仅测量位置或速度是不够的。

2.泽林格的研究组还包括下列人员：迪克·勃米斯特，潘建伟，克劳斯·马特尔，曼弗莱德·伊布与哈罗德·韦恩福尔特；德·玛蒂尼的研究组还包括：S.贾科米尼，G.米兰尼，F.西阿里诺以及E.罗姆巴蒂。

3.对那些熟悉量子力学体系的读者来说，这里是量子传输的关键步骤。假定我在纽约的那个光子的初始态具有如下形式$|\Psi>:=\alpha|0>:+\beta|1>:$，其中:0＞|与1＞为两个光子的极化态，我们再令系数正定，归一，但却可取任意值。我的目标是要给尼古拉斯足够的信息以便他能在伦敦制成一个处于完全相同的量子态的光子。为了达成这一目标，尼古拉斯与我首先获得一对处于纠缠态的光子，该纠缠态，比方说，可以为$|\Psi>_{23}=(1/\surd 2)0_2 0_3>-(1/\surd 2)1_2 1_3>$，因而，三光子系统的初始态就是$|\Psi>_{123}=(\alpha/\surd 2)\{|0_1 0_2 0_3>-|0_1 1_2 1_3>\}+(\beta/\surd 2)\{|1_1 0_2 0_3>-|1_1 1_2 1_3>\}$。当我对光子1和光子2进行贝尔态测量的时候，我就将这个态投射到下面4个态中的一个：$|\Phi>_{\pm}=(1/\surd 2)\{|0_1 0_2>\pm|1_1 1_2>\}$以及$|\Omega>_{\pm}=(1/\surd 2)\{|0_1 1_2>\pm|1_1 0_2>\}$。现在，如果我用粒子1和2的本征态为基重新表示初始态，则有：$|\Psi>_{123}=1/2\{|\Phi>_{+}(\alpha|0_3>-\beta|1_3>)+|\Phi>_{-}(\alpha|0_3>+\beta|1_3>)+|\Omega>_{+}(-\alpha|1_3>+\beta|0_3>)+|\Omega>_{-}(-\alpha|1_3>-\beta|0_3>)\}$。因而，在我测量之后，我使这个4种态叠加起来的系统“塌缩”到了其中的一种上。一旦我跟尼古拉斯通信（通过普通的办法）告知他我发现的是4种态中的哪一种，他就会知道该如何操作光子3使之被复制为初始的光子1。比方说，如果我发现测量结果是$|\Phi>_{-}$，那尼古拉斯就不需要对光子3采取任何行动，因为，如上所述，它就已经是光子1的初始态了。如果我得到的是其他结果，尼古拉斯就需要做一些适当的转动（如你所知，是根据我得到的结果进行操作），以

使光子3处于需要的态上。

4.事实上，数学比较好的读者将会注意到，证明所谓的量子不可克隆定理不算很难。假定我们有一个幺正克隆算符U，将任意态作为输入，这个算符可以输出两个同样的态（对于任意给定的|α＞，将U作用于|α＞→|α＞|α＞）。注意，将U作用于（|α＞＋|β＞）这样的态会得到（|α＞|α＞＋|β＞|β＞），而这并不是原始态的双重拷贝（|α＞＋|β＞）（|α＞＋|β＞）。因而，并不存在这样一个可以用于量子克隆的算符U。（伍特斯与祖莱克于20世纪80年代早期率先做出证明）。

5.参与到量子传输的理论与实验实现的发展中的研究人员还有很多。除了在正文中提到的那些，当时还在剑桥大学的佐藤胜彦也在罗马实验中扮演了重要角色，加州理工学院的杰弗瑞·金博尔研究组也曾在量子态连续性质的超距传输中取得领先。

6.要想对多粒子纠缠体系那极其有趣的进展有一番了解的话，可以参考一下诸如B.Julsgaard, A.Kozhekin以及E.S.Polzik合写的“Experimental long-lived entanglement of two macroscopic objects”Nature 413（SePt.2001），400-403。

7.利用量子纠缠与量子传输的众多研究领域中，最激动人心也最活跃的一个就是量子计算领域。要想对近年来的量子计算状况有一个了解，可以参考Tom Siegfried的The Bit and the Pendulum（New York：John Wiley，2000）以及George Johnson的著作A Shortcut Through Time（New York：Knopf，2003）。

8.速度增加时时间变慢的一个效应——我们在第3章没有讨论但是在本章却非常重要——是所谓的双生子佯谬。这一问题很容易描述：如果你我以匀速相对于彼此运动，我会认为你的时钟比我的慢一些。但是因为你与我都可以宣称自己处于静止系，所以在你看来是我

的时钟变慢了一些。我们两个都以为是对方的时钟变慢了，这里看起来存在着矛盾，但实际并非如此。匀速运动的时候，我们双方的时钟始终在远离彼此，因而没法面对面的比对一下，看看到底谁的时钟“真的”慢。而其他间接的比对（比方说，我们可以通过手机通信来对比一下时间）都会因为空间距离的存在而花一些时间，而这就必然会带来不同观测者对“此刻”定义不同的这种复杂性，如我们在第3章及第5章中讨论过的那样。在这里我不打算深究这个问题，总之要记住，一旦将这种狭义相对论所带来的复杂性处理清楚了，我们每个人都宣称对方的时钟慢于自己这件事就不再矛盾了（可以参考，诸如E.Taylor与J.A.Taylorde《时空物理》，以了解一下完整的、技术层面的同时也是基本意义上的讨论）。使事情变得更为复杂的是，你减速，停下，转弯，朝我过来面对面的比对时钟，以图消除不同的“此刻”定义所带来的复杂性。当我们面对面的时候，到底谁的时钟变慢了呢？这就是所谓的双生子佯谬：如果你和我是一对双胞胎，当我们再次碰面的时候，我们是一样大呢，还是我们中的某一个看起来老些呢？答案是我的时钟会比你的时钟走得更快——如果我们是双胞胎，我更老些。我们可以有很多种方式来解释为什么会这样，最简单的是要注意到，当你改变你的速度经历加速度时，我们在视角上的对称性消失了——你绝对可以宣称你在运动（因为你可以感觉到它——或者，用我们在第3章中的讨论，不同于我，你的旅程在时空中留下的不会是一条直线），因而你的时钟会慢于我的时钟，对你来说，时间流逝得更少。

9.约翰·惠勒曾提出过一种量子宇宙中以观测者为中心的可能性。归结到他著名的格言就是“在成为被观测到的现象之前，任何基本现象都不能被算作现象”。你可以在约翰·惠勒与肯尼斯·福特合著的Geons, Black Holes, and Quantum Foam：A Life in Physics（New York：Norton，1998）中对惠勒多姿多彩的物理人生有更多的了解。罗杰·彭罗斯也曾在他的著作The Emperor's New Mind以及Shadows of the Mind：A Search for the Missing Science of

Consciousness（Oxford：Oxford University Press，1994）探讨过量子物理与意识的关系。

10.可参见，例如，P.A.Schilpp编辑的The Library of Liuing Philosophers卷七《阿尔伯特·爱因斯坦》中的“Reply to Criticisms”（New York：MJF Books，2001）。

11.W.J.van Stockum, Proc.R.soc.Edin.A57（1937），135.

12.专业读者应该能看出来我在这里做了简化。1966年，约翰·惠勒的学生罗伯特·格罗克证明，不通过折叠空间的方法构建蠕虫洞至少在理论上是行得通的。但是不同于更为直观的，通过折叠空间的方法构建蠕虫洞——在这个方法中只有蠕虫洞还不能实现时空旅行，在格罗克的方法中，构建阶段本身就会要求时间发生扭曲，而人们可以自由的来往于古今（但不能回到构建事件开始之前）。

13.简略地讲，如果你以接近光速的速度穿越一块含有这些奇异物质的区域，并将你所测得的能量密度取平均，你会发现你得到了一个负值。按物理学家的说法，这样的奇异物质破坏所谓的平均弱能量条件。

14.实现奇异物质的最简单办法为第12章中讨论过的卡西米尔实验，所依靠的就是平行板间电磁场的真空涨落。计算表明，正是平行板间的量子涨落与真空中的量子涨落之间的差值，导致了负的平均能量密度（以及负压）。

15.如果只想在科普意义上而非技术层面上了解一下蠕虫洞，那么可以参考Matt Visser的Lorentzian Wormholes：From Einstein to Hawking（New York：American Institute of Physics Press，1996）。

第16章

1.数学比较好的读者可以回忆一下第6章注释6的内容。熵被定义为重排数目（或者说状态数目）的对数，而这一点对于给出本例的正确结果非常重要。你将两个特百惠家用塑料容器对接起来，空气分子的各种状态就可以这样描述：先给出第一个容器中的空气分子状态，再给出第二个容器中的空气分子状态，因而两个容器连接起来后的重拍数目就等于分开时每一个的重拍数目的平方。取对数之后，熵变为单独一个容器中的熵的2倍。

2.你应该注意到，拿体积和面积做比较没有任何意义，因为两者的单位根本就不同。正如正文中说明的那样，我在这个地方的真正意思是，体积随半径变大而变大的速率要快于面积随半径变大而变大的速率。因为熵正比于面积而不是体积，所以熵随着某一区域尺寸的变大而变大的速率就要小于若熵正比于体积的话的变大速率。

3.尽管这里已经抓住了熵界的本质，但是专业读者可能会注意到我做了简化。由拉斐尔·波索提出的更加准确的界，指出通过零超曲面（任何一点都有非正的聚焦参数Θ）的熵流被限定在A/4内，其中A为零超曲面的类空截面的面积（“光片”）。

4.更准确地说，一个黑洞的熵，等于普朗克单位下其视界的面积除以4，再乘以波尔兹曼常数后的结果。

5.数学比较好的读者可以回忆一下第8章的注释。还有另一种视界概念——宇宙视界，它是观测者能够以及不能够与之具有因果性联系的事物之间的分割面。人们相信对于这样的视界，其面积仍正比于熵。

6.1971年，匈牙利裔物理学家丹尼斯·盖博因全息照相术的发现而被授予诺贝尔奖。盖博从20世纪40年代开始一直致力于寻求从物体上反射回来的光波中捕获更多信息的方法，而其最初的动机是改进电子显微镜的分辨能力。我们以照相机为例来说明盖博的研究，照相机会

记录下从物体上反射回来的光波的强度；光强越高，相片上的相应位置就会越亮而光强越低，相片上相应位置就会越暗。盖博和另一些人认识到，光强只是光波所携带的部分信息。比如说我们观察图4.2（b）：尽管干涉图案受光的强度（振幅）影响（振幅更高的波产生更亮的图案），但是图案的产生却是因为，来自于每一个小缝的叠加波在沿着探测屏的方向上达到波峰、波谷或中间高度的位置不同。波的这后一种信息即是所谓的相位信息：若两列波在某一点彼此加强（两列波同时到达波峰或同时到达波谷），则称这两列波同相；若两列波在某一点彼此削弱（一列波到达波峰而另一列波到达波谷），则称这两列波异相，而且，更为普遍的是，这两列波有在这两种极端情况之间——部分加强，部分削弱——的相位关系。相干图案记录的就是相干光波的相位信息。

盖博发明了一种在特别设计的胶片上同时记录从物体上反射的光波的强度和相位的方法。用现代的语言说，他的方法非常类似于图7.1中的实验设置，只不过其中一束激光在射向探测屏时会被感兴趣的物体弹回。如果屏上备有包含了适当感光乳胶的底片，就将记录下自由传播的光束与被物体反射回来的光束之间的相干模式——以在胶片表面上留下微小的刻蚀线的方式记录。相干模式将会既记录下反射回来的光的强度，又记录下两束光之间的相位信息。盖博的思想为科学带来的后续影响非常重要，为范围广泛的多种测量技术带来了巨大的进步，但是对于普通大众来说，全息术对人们的影响更多地体现在艺术和商业领域。

普通的照片之所以看起来是平面的，在于其所记录的只是光强。要想得到深度的话，你还需要相位信息。原因在于，当光波传播的时候，它会周期性的经历波峰和波谷，所以相位信息——或者更准确地说，从物体上相邻点反射回来的光束的相位差——记录了光线传播远近的差别。比如说，当你从正面看一只猫时，猫的眼睛比猫的鼻子要远一点，这种远近之差就被记录在从猫脸部上不同部位反射回来的光

的相位差中。用一束激光照在全息图上，我们就能得到这张全息图所记录的相位信息，于是我们就可以将深度也添加到图片中。其结果我们都曾见过：二维塑料片上显现出了令人吃惊的三维影像。但需要注意的是，你的眼睛并不是用相位信息看出深度。相反的，你的眼睛用的是视差：从给定点发出，进入你的左眼以及右眼的光线在角度上的微小差别就是你得到的信息，大脑再将信息解码成点的远近。这就是——比方说——当你失去一只眼的视力时（或者闭上一只眼时），你对深浅远近的感知会受到影响的原因。

7.对更加偏爱数学语言的读者而言，这里可以这样表述：一束光，或者按更具普遍意义的说法一束无质量粒子，可以在有限时间内从反德西特空间的内部传播到无限远的空间然后返回。

8.对更加偏爱数学语言的读者，马达西纳的工作是在$Ad_{S5}XS_5$的框架下展开的，其边界理论来自于$AdS_5$的边界。

9.这一说法更像是社会学的语言而不是物理学的。弦论脱胎于量子粒子物理学，而圈量子引力则传承于广义相对论。但是我们必须注意到，就今天而言，只有弦论才能和广义相对论的成功预言联系起来，因为只有弦论才能在大距离尺度上令人信服的回归到广义相对论。圈量子引力虽然在量子范畴内很好理解，但其与大尺度现象之间的鸿沟却很难拉近。

10.更准确地说——正如我们在《宇宙的琴弦》一书中的第13章所讨论过的——自从贝肯斯坦和霍金在20世纪70年代做了哪些工作，我们已经知道了黑洞中有多少熵。但是，这些研究者们走过的途径却相当迂回，且从未能和用以说明他们所发现的熵的微观重排（参见第6章）很好的相容。20世纪90年代中期，两位弦论学家弥补了这一空白。安德鲁·斯特劳明格和卡姆兰·瓦法很巧妙地发现了黑洞和弦论或M理论中的膜的某些结构之间的关系。简单地说，他们能够证明，在膜的某些特别组合下，某些特别的黑洞能够容纳正好等于其基本组分

（不论其基本组分是什么）重排数的熵。他们在数这些膜的重排数（取对数）时发现所得到的答案正好是对应黑洞的表面积（普朗克单位下）除以4——正好就是多年前发现的黑洞的熵。在圈量子引力中，研究人员能够证明黑洞的熵正比于其表面积，但是得到精确的结果（普朗克单位下的表面积除以4）却绝非易事。如果某一参数，所谓的Immirzi参数，选择得当，那么准确的黑洞熵可以从圈量子引力的数学中得出；但从理论本身，人们却找不到能被普遍接受的基本解释，来说明这一参数的正确值究竟是怎样得到的。

11.在整个章节中，我都在压缩定量上很重要但是概念上没什么用处的数字参数。

# 术语表（以拼音为序）

D膜，狄利克雷P-膜（D-branes, Dirichlet-p-branes）：P-膜具有“黏性”开弦的端点可以附着在P-膜上。

M理论（M-theory）：将5种不同版本的弦论统一起来的理论，关于所有的力与所有的物质的完整的量子力学理论，目前尚不完备。

P膜（p-brane）：空间维度为P的弦论或M理论的要素之一。也见D膜。

W粒子，Z粒子（W and Z particles）：传递弱核力的信使粒子。

**A**

暗能量（dark energy）：假想中均一的充满于空间的能量或压力；是一个比宇宙常数更具普遍性意义的概念，因为其能量或压力随时间变化。

暗物质（dark matter）：充满空间的物质，有引力效应但是并不发光。

**B**

暴涨宇宙学（inflationary cosmology）：宇宙学理论，认为早期宇宙曾经有过一个空间迅速膨胀的短暂时期。

暴涨子场（inflaton field）：能量与负压驱动暴涨发生的场。

背景独立性（background independence）：在一个物理理论中，空间和时间来自于更基本的概念，而不是自明地存在于理论中。

闭弦（closed string）：弦论中的能量丝，环形。

标准蜡烛（standard candles）：内禀亮度已知的物体；可用于测量天文学距离。

标准模型（standard model）：由量子色动力学以及电弱理论组成的量子力学理论；描述除引力外的所有物质与力。基于点粒子概念。

波函数（wavefunction）：见概率波。

不确定原理（uncertainty principle）：量子力学原理；互补的两种物理性质同时测量时其不确定度要受到一个基本的限制。

**C**

场（field）：弥漫于空间的“迷雾”或“物质”；传递力或者描述粒子的存在、运动。数学上，在空间中的每一个点用一组量子数来表示场的值。

超对称（supersymmetry）：一种对称性；将整数自旋的粒子（传递力的粒子）与半整数自旋的粒子（物质粒子）交换时，物理定律保持不变的性质即为超对称性。

超弦理论（superstring theory）：基本元素为一维的圈（闭弦）或者振动能量片（开弦）的理论，这一理论将广义相对论与量子力学统一起来；并且引入了超对称。

**D**

大爆炸理论/标准大爆炸理论（big bang theory/standard big bang theory）：认为宇宙自其诞生后便处于膨胀的理论。

大收缩（big crunch）：一种可能的宇宙终结形式，类似于大爆炸的反面；空间向自身塌缩。

大统一（grand unification）：试图将强力、弱力，以及电磁统一在一个理论框架下的理论。

电磁场（electromagnetic field）：施加电磁力的场。

电磁力（electromagnetic force）：自然界4种基本力之一；作用于具有带电荷的粒子。

电弱理论（electroweak theory）：将电磁力与弱核力统一为电弱力的理论。

电弱希格斯场（electroweak Higgs field）：在冰冷虚无的空间中获得非零值的场；赋予基本粒子以质量。

电子场（electron field）：电子为最小的组分的场。

对称性（symmetry）：在某种变换下，物理系统保持不变的性质就是对称性（例如，将一个完美的球绕球心作旋转运动时，球的表面不发生变化）；对物理系统的变换不会对描述该物理系统的物理规律起作用，即称该物理规律具有此种变换对称性。

对称性自发破缺（sPontaneous symmetry breaking）：专业文献中对希格斯海形成的称法；对称性自发破缺意味着本来明显的对称性隐藏了起来或者被破坏了。

多世界诠释（Many Worlds interpretation）：一种量子力学解

释，所有概率波允许的可能性都能在不同的宇宙中得以实现。

**F**

负曲率（negative curvature）：小于临界密度的空间的形状，马鞍状。

**G**

概率波（probability wave）：量子力学中描述在给定位置发现粒子的概率的波。

概率波塌缩，波函数塌缩（collapse of probability wave, collapse of wavefunction）：概率波的变化，展开的概率波变成窄峰的形状。

干涉（interference）：叠加起来的波产生不同图样的现象。在量子力学里，干涉意味着看起来完全不同的可能性组合到一起。

哥本哈根诠释（Copenhagen interpretation）：一种对量子力学的诠释。大尺度上的物体遵从经典物理定律，而小尺度上的物体遵从量子力学定律。

惯性（inertia）：物体保持运动状态不变的性质。

光以太（luminiferous aether）：见以太。

光子（Photon）：电磁力的信使粒子，一“束”光。

广义相对论（general relativity）：爱因斯坦的引力理论；空间与时间是弯曲的。

**H**

黑洞（black　hole）：一种物理实体；只要距离足够近（小于黑洞视界），黑洞巨大的引力场就可以吞噬一切，即使光也不能例外。

**J**

加速度（acceleration）：速度在大小与（或）方向上的改变。

加速器，原子碰撞机（accelerator, atom　smasher）：粒子物理的研究工具；将粒子以高速对撞。

胶子（gluon）：强核力的信使粒子。

经典物理（classical　physics）：本书中指的是牛顿理论与麦克斯韦理论；一般指的是所有的非量子力学理论，包括狭义相对论与广义相对论。

纠缠，量子纠缠（entanglement, quantum　entanglement）：一种量子现象，空间上相隔很远的两个粒子之间具有关联性。

绝对空间（absolute　space）：牛顿的空间观；将空间视作永不改变的客体，并且与其内部的一切事物相独立。

绝对论者（absolutist）：持绝对空间观点的人。

绝对时空（absolute　spacetime）：狭义相对论的空间观；将空间与时间视作不可分割的统一整体，这个统一整体不会发生改变且独立于其内部的一切事物。

**K**

卡鲁扎克莱因理论（Kaluza-Klein　theory）：物理理论，这一理论中的宇宙可以具有高于3的空间维度。

卡西米尔力（Casimir force）：真空场涨落的不平衡产生的量子力学力。

开尔文（Kelvin）：利用绝对零度（最低温度，在摄氏温度制中为-273℃）来标度的温度单位。

开弦（open string）：弦论中的能量丝，像一个小片。

可观测宇宙（observable universe）：在我们的宇宙视界之内的那部分宇宙；这部分宇宙由于距离我们足够近，现今的我们可以观测到它所发出的光，也就是说这部分是我们能够看到的宇宙。

夸克（quarks）：受强核力支配的基本粒子；有6种类型的夸克（上夸克，下夸克，奇异夸克，粲夸克，顶夸克与底夸克）。

**L**

量子测量疑难（quantum measurement problem）：概率波所具有的无数可能性如何在测量的时候让位于唯一的结果，这一问题就是量子测量疑难。

量子力学（quantum mechanics）：20世纪20～30年代建立的理论，描述原子及亚原子尺度的物理。

量子色动力学（quantum chromodynamics）：强核力的量子力学理论。

量子涨落（quantum fluctuations, quantum jitters）：由于不确定原理的存在，场的值在小尺度上不可避免的急速变化。

临界密度（critical density）：使空间保持平坦所需的质量或能量密度；大约每立方米$10^{-23}$克。

路径选择信息（which-path information）：声明粒子从源到探测器所走路径的量子力学信息。

## M

马赫原理（Mach's principle）：马赫提出的原理，所有的运动都是相对的，静止的标准由宇宙中的平均质量分布提供。

膜世界方案（braneworld scenario）：弦论或M理论所引申出来的一种可能性，我们熟悉的三维世界实际上是一张3-膜。

## N

能量碗（energy bowl）：见势能碗。

## P

平坦空间（flat space）：一种可能的宇宙空间形态，其中没有弯曲。

平移不变性，平移对称性（translat ionalinvariance, translational symmetry）：已知自然定律的一种性质；即在不同空间方位处的自然定律都是一样的。

平直性疑难（flatness problem）：宇宙学理论必须回答的疑难，即所观测到的空间是平直的。

普朗克长度（Planck length）：在这一长度（10-33厘米）之下，量子力学和广义相对论之间的矛盾变得明显；传统的空间概念不再适用。

普朗克时间（Planck time）：指10-43秒长的时间，这一时间内，光可以传播普朗克长度的距离；在这样小的时间间隔内，传统的

时间概念不再成立。

普朗克质量（Planck mass）：指10-5克大小的质量（这样大的质量只相当于一点灰尘的重量，但却是质子质量的一千亿亿倍）；弦的特征质量。

**Q**

强核力（strong nuclear force）：对夸克起作用的自然力；将所有的夸克封闭于质子和中子之内。

**R**

热力学第二定律（second law of thermodynamics）：平均起来，任意时刻物理系统的熵总是趋向于增加的方向。

弱核力（weak nuclear force）：自然界中的基本力之一，在亚原子尺度起作用；诸如放射性之类的现象即与其有关。

**S**

熵（entropy）：物理系统混乱度的量度；大小与系统所有基本状态数有关。

时间反演对称性（time-reversal symmetry）：已知自然定律的一种性质，即时间指向不同方向时物理定律不发生变化。在任意时刻，过去的物理定律都与未来的物理定律精确一致。

时间片（timeslice）：同一时刻的所有空间；整块时空的一片。

时间之箭（arrow of time）：时间指向——从过去到未来。

时空（sPacetime）：将时间与空间整合到一起的概念，由狭义

相对论首先引出的概念。

势能（Potential energy）：场或物体中的能量。

势能碗（Potential energy bowl）：场所具有的能量在给定值的形状，专业文献中称为场的势能。

视界（eventhorizon）：黑洞外的假想球面；进入球面内的任何物质都无法逃离黑洞的引力。

视界疑难（horizon Problem）：宇宙学理论面对的一大疑难，即解释在彼此宇宙学视界之外的不同空间区域为何具有类似的性质。

速度（velocity）：物体运动的大小和方向。

**T**

统一理论（unified theory）：将所有的力与物质纳入同一框架的理论。

**W**

微波背景辐射（microwave background radiation）：见宇宙微波背景辐射。

**X**

希格斯场（Higgs field）：见电弱希格斯场。

希格斯场真空期望值（Higgs field vacuum expectation value）：希格斯场在真空中获得非零期望值；希格斯海。

希格斯海（Higgso cean）：在本书中特指希格斯场真空期望值。

希格斯粒子（Higgs particles）：希格斯场的量子。

狭义相对论（special relativity）：爱因斯坦的理论，时间和空间并不是单独绝对的，而是取决于不同观测者的相对运动。

弦论（string theory）：物理理论，基本研究对象为振动着的一维能量丝（见超弦理论）；不需要在理论中引入超对称。有时也用作超弦理论的简称。

相变（phase transition）：温度在足够大的范围内变化时，物理系统会发生性质上的突变。

相对论者（relationist）：持如下观点的人：所有的运动都是相对的，空间不是绝对的。

信使粒子（messenger particle）：最小单位的力，力的效应通过其得以传递。

**Y**

以太，光以太（aether, luminiferou saether）：假想中占据空间的物质，光以其为传播介质；并不存在。

引力子（gravitons）：假想的传递引力的信使粒子。

宇宙常数（cosmological constant）：假想中均匀地充满于空间的能量或压力；起源与组成未知。

宇宙视界，视界（cosmic horizon, horizon）：宇宙开始之后，其外的光就无法到达我们的位置。

宇宙微波背景辐射（cosmic microwave background radiation）：早期宇宙留下的剩余电磁辐射，弥漫于整个空间。

宇宙学（cosmology）：研究宇宙起源及演化的学科。

**Z**

真空（vacuum）：最虚无的空间；最低能量态。

转动不变性，转动对称性（rotational invariance, rotational symmetry）：物理系统或物理定律在转动变换下保持不变的性质。

自旋（sPin）：基本粒子的量子力学性质。有点类似于陀螺的转动，基本粒子也有转动（具有内禀角动量）。

# 推荐书目

有关空间与时间这一问题，无论是一般性文献还是技术性文献数目都非常巨大。下面所列出的这些参考文献大都适合一般读者阅读，但其中也有少量文献要求读者有一定的知识储备。就我个人而言，这些文献都是大有裨益的，相信对于那些渴望进一步探索本书中所提到的各种理论进展的读者也能起到很好的引路作用[97]。


Albert, David. Quantum Mechanics and Experience. Cambridge, Mass.：Harvard University Press，1994.

——. Time and Chance.Cambridge, Mass.：Harvard University Press，2000.

Alexander, H. G.The Leibniz-Clarke Correspondence. Manchester, Eng.：Manchester Uni-versity Press，1956.

Barbour, Julian. The End of Time.Oxford：Oxford University Press，2000.

——and Herbert Pfister. Mach's principle. Boston：Birkh?user，1995.

Barrow, John. The Book of Nothing.New York：Pantheon，2000.

Bartusiak, Marcia. Einstein's Unfinished Symphony. Washington, D.C.：Joseph Henry Press，2000.

Bell, John. Speakable and Unspeakable in Quantum Mechanics. Cambridge, Eng.：Cam-bridge University Press，1993.

Blanchard, Ph.，and D. Giulini, E.Joos, C.Kiefer, I.-O Stamatescu. Decoherence：Theoretical, Experimental and Conceptual Problems. Berlin：Springer，2000.

Callender, Craig, and Nick Hugget. Physics Meets Philosophy at the Planck scale.Cam-bridge, Eng.：Cambridge University Press，2001.

Cole, K. C.The Holein the Uniuerse.New York：Harcourt，2001.

Crease, Robert, and Charles Mann. Thesecond Creation.New Brunswick, N.J.：Rutgers University Press，1996.

Davies, Paul. About Time.New York：Simon & Schuster，1995.

——. Howto Build a Time Machine.New York：Allen Lane，2001.

——. Spaceand Time in the Modern Uniuerse. Cambridge, Eng.：Cambridge University Press，1977.

D'Espagnat, Bernard. Veiled Reality. Reading, Mass.：Addison-Wesley，1995.

Deutsch, David. The Fabric of Reality.New York：Allen Lane，1997.

Ferris, Timothy. Coming of Age in the Milky Way.New York：

Anchor，1989.

——. The Who leshebang.New York：Simon & Schuster，1997.

Feynman Richard. QED.Princeton：Princeton University Press，1985.费曼，《QED光与物质的奇异性》，商务印书馆。

Fölsing, Albrecht.Albert Einstein.New York：Viking，1997.

Gell-Mann, Murray. The Quarkand the Jaguar.New York：W.H.Free man，1994.盖尔曼，《夸克与美洲豹：简单性和复杂性的奇遇》，湖南科学技术出版社。

Gleick, James. Isaac Newton.New York：Pantheon，2003.

Gott, J. Richard.Time Trauel in Einstein's Uniuerse. Boston：Houghton Mifflin，2001.

Guth, Alan. The In flationary Uniuerse. Reading, Mass.：Perseus，1997.

Greene, Brian. The Elegant Uniuerse.New York：Vintage，2000.布赖恩·格林，《宇宙的琴弦》，湖南科学技术出版社。

Gribbin, John. schr?dinger's Kittens and the search for Reality. Boston：Little, Brown，1995.

Hall, A. RuPert. Isaac Newton.Cambridge, Eng.：Cambridge University Press，1992.

Halliwell, J. J.，J.Pérez-Mercader, and W.H.Zurek.Physical Origins of Time Asymmetry. Cambridge, Eng.：Cambridge University Press，1994.

Haw king, Ste Phen. The Uniuersein a Nutshell.New York：Bantam，2001.霍金，《果壳中的宇宙》，湖南科学技术出版社。

——and Roger Penrose. The Nature of Space and Time.Princeton：Princeton University Press，1996.霍金，罗杰·彭罗斯，《时空本性》，湖南科学技术出版社。

——，KiP Thorne, Igor Novikov, Timothy Ferris, and Alan Lightman. The Future of sPace-time.New York：Norton，2002.

Jammer, MaX. Conce Ptsofs Pace.New York：Dover，1993.

Johnson, George. Ashortcut Through Time.New York：Knopf，2003.

Kaku, Michio. HyPersPace. New York：Oxford University Press，1994.

Kirschner, Robert. The Extrauagant Uniuerse. Princeton：Princeton University Press，2002.

Krauss, Lawrence. Quintessence. New York：Perseus，2000.

Lindley, David. Boltzmann's Atom.New York：Free Press，2001.

——. Where Does the WeirdnessGo？ New York：Basic Books，1996.

Mach, Ernst. The science of Mechanics.LaSalle, Ill.：OPen Court，1989.

Maudlin, Tim. Quantum Non-locality and Relatiuity. Malden,

Mass.：Blackwell，2002.

Mermin, N，David. Boojums All the Way Through.New York：Cambridge University Press，1990.

Overbye, Dennis. Lonely Hearts of the Cosmos.New York：Harper Collins，1991.

Pais, Abraham. subtle Is the Lord.Oxford：Oxford University Press，1982.亚伯拉罕·派斯，《爱因斯坦传》（上·下），时代文艺出版社。

Penrose, Roger. The EmPeror's New Mind.New York：OXford University Press，1989.罗杰·彭罗斯，《皇帝新脑》，湖南科学技术出版社。

Price, Huw. Time's Arrow and Archimedes' Point.New York：Oxford University Press，1996.

Rees, Martin. Beforethe Beginning.Reading, Mass.：Addison-Wesley，1997.

——. Justsix Numbers.New York：Basic Books，2001.马丁·里斯，《6个数：塑造宇宙的深层力》，上海科学技术出版社。

Reichenbach, Hans. The Direction of Time.Mineola, N.Y.：Dover，1956.

——. The Philosophy of Spaceand Time.New York：Dover，1958.

Savitt, Steven. Time'sArrows Today. Cambridge, Eng.：

Cambridge University Press，2000.

Schr?dinger, Erwin.What Is Life？ Cambridge, Eng.：Canto，2000.欧文·薛定谔，《生命是什么》，湖南科学技术出版社。

Siegfried, Tom. TheBitand the Pendulum.New York：John Wiley，2000.

Sklar, Lawrence. sPace, Time, and sPacetime.Berkeley：University of California Press，1977.

Smolin, Lee. Three Roadsto Quantum Grauity.New York：Basic Books，2001.

Stenger, Victor. Timeless Reality. Amherst, New York：Prometheus Books，2000.

Thorne, KiP. Black Holesand Time WarPs.New York：W.W.Norton，1994.基普·S.索恩，《黑洞与时间弯曲（爱因斯坦的幽灵）》，湖南科学技术出版社。

von Weizs?cker, Carl Friedrich.The Unity of Nature.New York：Farrar, Straus, and GirouX，1980.

Weinberg, Steven. Dreams of a Final Theory.New York：Pantheon，1992.温伯格，《终极理论之梦》，湖南科学技术出版社。

——. The First Three Minutes.New York：BasicBooks，1993.史蒂文·温伯格，《宇宙最初三分钟——关于宇宙起源的现代观点》，中国对外翻译出版公司。

Wilczek, Frank, and Betsy Devine. Longing forthe Harmonies.New York：Norton，1988.

Zeh, H. D.The Physical Basisofthe Direction ofTime.Berlin：SPringer，2001.

[1]译注：《格列佛游记》的主人公。

[2]译注：巴特·辛普森为美国流行动画片《辛普森一家》中的人物，他是居住在斯普林菲尔德的辛普森家中的长子，性格调皮捣蛋，不擅读书学习。他的父亲霍默·辛普森也将在下文出现。下文中出现的巴特·辛普森和霍默·辛普森以及斯普林菲尔德我们将不再一一提及。

[3]虽然离心力和向心力这样的词语常常被用来描述旋转运动，但是它们只是一个名称而已。我们真正想知道的是为什么旋转会产生力。

[4]关于在下文中提到的马赫观点有些需要交代的地方。马赫的某些著作含混不清，而我们提到的某些观点实际上来自另外一些人对马赫工作的解释。一般认为马赫知道有这些解释，考虑到他并未对这些解释提出批评意见，所以人们通常认为马赫本人是同意这些解释的。为了还历史以其本来面貌，我每次写出“马赫认为”或者“马赫的想法”时，亲爱的读者，你应当将其视作“马赫提出的一种方法盛行的解释”。

[5]我之所以喜欢用人来举例子是因为这样能够一下子把我们正在讨论的物理问题与我们的内在感受联系起来。但是这也带来一个问题，那就是我们的全身都可以随意识活动，所以身体的某个部分可以相对于另外的部分运动——这样带来的效果是我们可以用身体的一部分作为基准来讨论另一部分的运动（比如你可以相对你的头旋转你的手臂）。所以我要强调统一的旋转运动——全身的每个部位都一起运动——来避免不必要的复杂性。所以，当我说到你的身体处于转动状态时，把你的身体想象成像牛顿绑在一起的石头或者奥运赛场上做出最后一个动作后保持静止的溜冰运动员一样。你全身的每个部位都以相同的速率转动。

[6]译注：拉斯塔法里教派，宣扬牙买加政治家加尔维（M.Garvey）的主张，认为黑人要反压迫必须回到非洲。长发是该教的一个标志。

[7]译注：《傻猫和坏鼠》为美国流行动画片《辛普森一家》中的风靡于少年儿童中的卡通片，其中主角为傻猫和坏鼠。

[8]像翻页相册的页数一样，图3.3中的页数只显示了代表性时刻。这就提示我们时间是否可以分离或是无限分割的。稍后我们将回到这个问题，但现在先想象一下时间是可以无限分割的，这样我们的翻页相册就可以在图中显示的时间片中插入无数页。

[9]画出蜷曲的空间很容易，但由于它们之间的联系紧密，时间也会因物质和能量而蜷曲。就像空间的蜷曲意味着空间被拉伸或压缩，如图3.10所示；时间的蜷曲也意味着时间可以被拉伸或压缩。也就是说，不同引力场中的时钟——比如一个在太阳上，另一个在外太空中——会以不同的速度运转。事实上，普通物体，比如地球和太阳（黑

洞不在此列），所造成的空间弯曲远不如它们所造成的时间弯曲明显。

[10]这种观点在狭义相对论中——广义相对论中引力场为零的特例——广为适用：零引力场仍是一种场，它可以测量且可以发生变化，因此为可定义的加速运动提供了一种标准。

[11]译注：史考莉和穆德都是美国流行剧《X档案》中的探员。

[12]译注：法国南部省份，下文中的爱克斯为该省的一个旅游胜地。

[13]译注：两者都是好莱坞著名影星。

[14]为避免语言上的复杂性，我把电子的自旋描述为关联，虽然更为正式的表述方式应为反关联：不管一个探测器发现什么样的结果，另一个都将发现相反的结果。为了与正式表述相比较，可以这样想象，我把其中一个探测器上的顺时针和逆时针的标签互换了。

[15]许多研究者，包括我在内，都相信贝尔的论证和埃斯拜科特的实验令人信服的解释了空间上相隔很远的粒子之间的可观测的关联性不能由史考莉式的论证说明（在他们的论证中，粒子的关联性源于粒子以前在一起时就具有的明确的、彼此关联的性质）。有一些人则试图避免或削弱由此而带来的、令人吃惊的非定域性结论。我并不赞成他们的怀疑，但我还是在本书最后注释部分列出了适宜于普通读者阅读的有关这些想法的读物。

[16]译注：圣·奥古斯丁（Saint Augustine，354～430），古代基督教主要作家之一，他认为时间是主观的，“存在于我们心中”。他对时间的哲学研究可参见其著作《忏悔录》；美国已故大法官波特·斯图尔特（Potter Stewart）曾这样描述色情业：“我无法给它下定义，但是我看一眼就知道是怎么回事（Iknowitwhen Iseeit）。”本书作者在这里只是借用一下这句名言。

[17]译注：亚历山大大帝的老师是著名的亚里士多德；美国独立战争期间，1776年12月25日，华盛顿带领军队横渡特拉华河。这次针对黑森雇佣兵的突袭行动是特伦顿战役的第一步。

[18]译注：埃尔维斯·普莱斯利（ElvisPresley，1935～1977），猫王，是20世纪美国最有影响的歌手之一。

[19]译注：大峡谷（Grand Canyon），美国亚利桑那州西北部高原由科罗拉多河切成的巨大峡谷，最宽的地方有29千米。光速30万千米/秒，所以差不多需万分之一秒才能穿过大峡谷。

[20]译注：丘巴卡，ChewBacca，昵称Chewie，电影《星球大战》中的人物。

[21]译注：动漫游戏《剑风传奇》中的名字。

[22]译注：协和式超音速客机，Concord aircraft，英国和法国联合研制的一种超音速客机，这种飞机一共只建造了20架。它的最大飞行速度可达2.04马赫。

[23]译注：1865年4月14日，约翰·维尔克斯·布思在福特大剧院刺杀了当时的美国总统林肯。

[24]在面包条上任选一点。取一片包含了该点的切片，使其与我们的现在时间片相交的角度小于45度。这个切片将代表一位远方的观测者的现在时间片——他对实在性的认识，开始时他相对于我们静止，就像丘巴卡，但现在以小于光速的速度相对于我们运动。而且我们可以使该切片包含你碰巧选取的面包中的那点（可以为任意的点）。

[25]也有不同于此处表述的例外，该例外与一些奇特的粒子有关。这与本章中讨论的问题可能没有多大关系，因此就不再进一步探讨了。如果你感兴趣，在注释2中有简要的讨论。

[26]注意时间反演对称性并不是指时间本身可以反过来或逆向“奔跑”。相反，正如我们所描述的，时间反演对称性是指按某种特殊的时间顺序发生的事件也可以按相反的顺序发生。更确切的描述可能是，事件反演或过程反演，又或事件顺序反演，但我们仍然沿用传统的说法。

[27]译注：CNN原指美国有线电视新闻网。作者这里自编了一个CelestialNews Network。

[28]熵是另一个术语使思想复杂化的例子。要是你把低熵意味着高度有序和高熵意味着低度有序（或者说高度无序）不停地弄混的话，那么别担心，我也常弄混。

[29]译注：美国著名电影导演。

[30]还记得吗？在前面的小节中，我们展示了一本仅有693页的双面印刷的书所能有的有序状态和无序状态在数目上的巨大差别。而我们现在要讨论的是差不多$10^{24}$个$H_2O$分子的情形，毫无疑问，其有序状态和无序状态在数目上的差别将是难以估量的。而且，对你以及环境（大脑、安全摄像机、气体分子，等等）中的所有原子和分子来说，同样的论证一样适用。也就是说，按照标准解释——靠着这个解释你可以相信自己的记忆——晚上10：30时，不但部分融化的冰块会处于更为有序——也就是更不可能——的状态，周遭的一切也会如此：摄像机记录下一系列事件时，熵会净增长（记录过程会释放出热及噪声）；类似的，事情在大脑中留下记忆时，尽管我们还没办法清楚地知道细节，但是，熵肯定会净增长（大脑的确会获得某种有序度的增加，但就

像任何会产生有序的过程一样，如果我们将所产生的热量也考虑进去，则熵会净增加）。因而，如果我们比较两套方案——一种是你相信自己的记忆，另一种则是，事物会自发的调整自己，使之从初始时的无序态自然地过渡到你现在，即晚上10：30，所看到的情形——之下，晚上10：00和10：30这段时间内酒吧中总的熵，我们就会发现，两种方案里的熵，差别非常之大。后一种方案，其方方面面都比前一种方案有更多的熵，而且是非常的多，因而，从概率的角度看，后一种方案更不可能发生。

[31]与之密切相关的一点是，要是我们使自己相信，我们此刻看到的世界刚刚从总的混沌中脱身出来，那么，完全相同的理由——需要使用之后的时刻——就会要求我们放弃我们所相信的当下的一切，转而将有序的世界归结为更为晚近的涨落。因而，按这种方式思考的话，每个下一时刻都会使相应的上一时刻时所相信的一切变得无效，这显然不是什么令人信服的解释宇宙之道。

[32]也就是说，给定大小的黑洞中所包含的熵比任何其他同等大小的物体中的熵都要多。

[33]译注：奥兹国，美国童话小说《绿野仙踪》中的国名，下面的多萝西和稻草人是其中的人物。

[34]尽管费曼的历史求和方法看起来突出了粒子方面，但它实际只是概率波（因为这种方法与每个粒子的所有历史有关，而每一种历史都有自己的概率贡献）的一种特殊诠释，可被纳入互补性的波动一面。而当我们说某物表现出粒子性时，我们指的其实是传统意义上的粒子会按单一轨迹运动这件事。

[35]如果你觉得这节有点难度，那么没关系，直接跳到下一节好了，不会连不上。但我鼓励你尝试一下，因为结果会对得起你的努力。

[36]公平地说，量子力学给人的印象并非是平滑渐变的，正如我们将在下一章中看到的那样，它所展现的是狂暴混乱的微观宇宙。这种混乱的起源就是波函数的概率性——即便某物此刻出现在这里，下一时刻它会有一定概率出现在全然不同的另一位置——而不是说波函数本身是一个变化无常的量。

[37]译注：斯波尔丁·格雷（SPaldingGray），美国演员，剧作家，表演艺术家。

[38]译注：雅加达，印度尼西亚首都；阿德莱德，澳大利亚港市；安克雷奇，美国阿拉斯加州南部的港口城市。

[39]若不用这个两维的气球模型作类比，而是直接用一个球形的三维空间作类比的话，在数学上当然非常简单，但即便是专业的数学家和物理学家也很难形象的勾画出这一图像。你倒也可以试着想象一个实心的三维球，比方说没有洞的保龄球。但这实

际上并不是一个可接受的形状。在我们想要的模型中，所有的点应该处于完全相同的地位，因为我们相信宇宙中所有的位置彼此类似（当然是在平均意义上）。但是保龄球上的点却彼此不同：有的就在表面，有的在内部，还有的正好位于球心位置。正如在两维气球模型中两维的表面围绕在三维的球体区域（包括气球中的空气）外，可接受的三维圆形也应该围绕在四维球体区域外。所以一个四维空间中的球体的三维表面才是一个可接受的形状。不过要是这么说还不能让你停止对一幅图像的渴求，那么你就可以像所有的专家做的那样：坚持利用易于想象的低维类比。事实上低维的类比几乎能够捕捉到所有本质特点。要不我们也可以考虑三维平直空间，与球体的圆形不同，平直空间具有可视化的特点。

[40]根据宇宙膨胀的速度是加速还是减速的不同，这样的星系所辐射出来的光可能会陷入令芝诺骄傲的困境之中：从另外一个星系辐射出来的光以光速穿越空间向我们飞来，但是我们两个星系之间的距离却以大于光速的速度扩大，从而导致向我们飞来的光永远都无法到达我们的身边。更为详尽的讨论可参见注释10。

[41]译注：电动兔，美国Energizer公司于1989年推出的一款玩具，打着鼓的兔子会一直向前走。

[42]电子游戏屏幕可以作为无边界的平直空间的有限大小版，没有边界的马鞍面也有其有限大小的现实例子。我不想在这里继续讨论这一问题，大家只需要知道所有这3种可能的曲率（正，零，负）都可以用无边界的有限大小的形状代表（理论上，宇航员版的麦哲伦可以在任意一种曲率的宇宙中实现太空大航海）。

[43]时至今日，宇宙中的物质远多于辐射，所以用与质量相关的单位——克每立方米——表示临界密度最方便。还要注意的是虽然临界密度10-23非常的小，但是宇宙中实在有太多的立方米了。而且，探讨的宇宙越古老，空间也就越小，能量或质量被压缩得也就越厉害，宇宙也就越质密。

[44]尽管对称性的减少意味着能够保持一切不变的操作更加少了，但在这些过程中释放到环境中的热却使得整体的熵——包括外部环境的熵——增加了。

[45]这一术语并没什么特别的重要之处，但是我们简要地看看为什么是这样的术语。图9.1（c）与图9.1（d）中的谷底具有对称的形状——圆环状——每一点都与其他点一样（每一个点代表的都是最低能量的希格斯场值）。但是，当希格斯场值滑到碗底的时候，它会落在圆环形谷底的某个特殊位置，这样一来就在谷底“自发”地选择了一个特殊位置。换句话说，这样一来，谷底上的点就不再具有相同的地位了，因为有一个点被挑了出来，所以希格斯场破坏了或者说“破缺”了这些点之间先前的对称性。因而，将这些话组合起来就是，希格斯滑落到谷底某个特定的非零值的这一过程被称为对称性自发破缺。我们将在正文中讲讲与这种希格斯海的形成有关的对称性减少的某些切实

方面。

[46]译注：引力特指我们常说的万有引力，也就是重力，是自然界中4种基本力中的一种。下文出现的吸引力或排斥力指的只是某种力具有吸引的性质还是排斥的性质。注意不要混淆。

[47]译注：《盖里甘的岛》，20世纪60年代的美国电影，讲述被遗弃在岛上的演员，开始时感觉愉快，但在他们等待救援的漫长时间里，一个个开始变得性格乖戾。

[48]物理学家们常常在某一名称后面加个“子”，表示相应的粒子。比如光加上“子”就是“光子”，电磁场的量子。

[49]译注：Secretariat，秘书，世界上跑得最快的纯种马，1973年在美国肯塔基赛马会上创下了30多年不破的世界纪录。

[50]译注：美国中部的平原州，面积20万平方千米。

[51]译注：Beatles的一首歌曲，名字就叫做《缀满钻石的天空中的露西》（Lucy In The sky With Diamonds）。

[52]橡胶带这个例子虽然很方便地说明问题，但是有其不完美之处。橡胶带所施加的向内的负压阻碍了盒子的膨胀；而暴涨子场的负压则推动了宇宙的膨胀。两者之间的这一重要区别正好折射出了我们在上一章的第一节末尾强调过的问题：在宇宙学中，驱动空间膨胀的并不是均匀的负压强（只有压强差才会产生力的效果；均匀的压强，无论正负，都不会产生力的效果）。压强和质量一样，都会贡献出引力。负的压强贡献出的是排斥性的万有引力，驱动空间膨胀的正是这种排斥性的万有引力。不过这一点并不影响我们的结论。

[53]宇宙膨胀的时候，光子的能量损失可因为其波长的变长——所谓的红移——而得以直接观测到，光子的波长越长，其能量损失也就越大。微波背景中的光子就经历了差不多140亿年的此类红移，而这正好解释了它们的很长的——还是微波——波长和低温。物质经历与此类似的动能（由于粒子运动而具有的能量）损失，但是束缚于粒子质量中的总能量（其静止能量——等于粒子静止时所具有的能量）并不变化。

[54]艾伦·古斯和艾迪·法依等多位研究人员已经探讨过在实验室中通过合成一小块暴涨子场从而创造一个宇宙的可能性。其中的麻烦之处并不仅仅在于我们还没能在实验上发现暴涨子场这一事实，还在于我们很难将重20磅的暴涨子场填充到边长不足$10^{-26}$厘米的空间内，要知道这样导致的密度将非常巨大——大约是原子核密度的1067倍——这一水平已经远超我们现有的实验能力。我们不但现在做不到这样的事情，很可能在将来也做不到。

[55]不要在这里搞糊涂了：我们在上一节讨论过的量子涨落的暴涨放大效应仍然会导致空间具有不均匀性，只不过这种不均匀性大概只有十万分之一那么点。这样小的不均匀性不会消除宇宙整体上的光滑性。我们现在讨论的是宇宙的这种整体上的光滑性是怎么出现的。

[56]译注：迈达斯（Midas）是希腊神话中小亚细亚中西部古国佛里吉亚（Phry-gia）国王，爱财，能点物成金。

[57]为了写作的便利，我们只讨论那些场强为零即达到其最低能量的场。对于其他场——比如希格斯场——的讨论类似于此，只是有一点不同，就是涨落围绕着场强的非零值波动，但场却具有最低能量。如果你想说，只有其中不包含任何物质且所有的场都不存在，而不是仅仅其场强为零的空间区域才算是真空的话，那么请先阅读一下本章节的注释。

[58]本章的剩余部分讲述了超弦理论的发现并讨论了与统一和时空结构有关的关键思想。读过《宇宙的琴弦》（特别是第6章和第8章）的读者会非常熟悉某些内容，对于这些读者来说，略掉本章的剩余内容继续下面章节的阅读毫无问题。

[59]还记得吗？在第9章的注释中我们曾经提过，即使微不足道的磁体所产生的磁力也会大于整个地球所带来的引力，从而吸起一个纸夹。数值上来说，引力大约是电磁力的$10^{-42}$倍。

[60]事实上还有另外一种合并广义相对论与量子力学的方法——圈量子引力，我们会在第16章中简要讨论一下这个问题。其支持者的观点更加接近前一种假设——时空在小尺度上具有不连续的结构。

[61]来自于希格斯海的质量与弦的振动的关系将在本章的后面内容中加以讨论。

[62]译注：罗杰·班尼斯特，生于1929年，英国人，1954年他成为第一个在4分钟内跑完1英里的人。

[63]译注：艺人，曾于1974年8月7日在世贸中心双塔间表演空中走钢索，花了一个小时跨越世贸中心两栋塔楼，之后遭逮捕入狱。后来，菲利普只能在中央公园为小朋友表演。

[64]译注：洛子峰，海拔8516米，为世界第四高峰，地处珠穆朗玛峰以南3千米处，它们之间隔着一条山坳，即通常说的“南坳”。

[65]如果你非要把左和右、顺时针和逆时针分开来算，那么你会认为蚯蚓有四个方向可以选择。但是我们说“独立”这个词的时候，我们将那些在同一几何轴上的方向——比如左和右，又比如顺时针和逆时针都是在同一个几何轴上——都算作一个方向。

[66]现在我们来为下面一章中将会遇到的内容做些准备，以便你能很好地了解相关进展。弦论学家早在几十年前就清楚地知道他们通常在弦论中用于数学分析的方程实际上只是某种近似（严格的方程早就被证明难于分析及理解）。不过，大部分人都认为近似方程已经精确到可以确定所需的额外空间维数的程度。近年来（令本领域内的很多物理学家非常感到震惊），一些弦论学家证明近似方程实际上丢掉了一维；现在人们普遍认为弦论需要7个额外的空间维度。我们将会看到，这并不会危及本章中讨论的内容，只是告诉我们本章中的内容也适用于一个更大的实际上更具统一性的理论框架。20

[67]译注：大号，大型带活塞的低音铜管乐器。

[68]译注：维克罗Velcro牌的搭扣，由钩和毛两种结构组成，看看你的衣服上有没有这种扣子。

[69]这些黏黏的家伙准确的名字是狄利克雷P-膜，或者简称为D-P-膜。我们将它简写为P-膜。

[70]译注：菲利普·佩蒂特（PhiliP PePetit），法国杂技名人。1974年，佩蒂特在纽约世贸大楼间架起钢丝，并在钢丝上行走、舞蹈了45分钟。

[71]哈佛大学的丽莎·兰德尔和约翰·霍普金斯大学的拉曼·山德拉姆提出了另一种方案。在他们的理论中，引力也有可能被限制住，但却不是被限制在膜上，而是通过额外维度的蜷曲方式将引力限制住，他们的方案使得额外维度的尺寸所受到的限制变得更加宽松。

[72]译注：弗兰克·西纳特拉，美国著名艺人，成名于纽约。

[73]译注：美国著名节目主持人，7次婚姻。

[74]译注：Trinity，圣经中三位一体的意思。建造世界上第一颗原子弹的实验室选址于此后，其项目领导人奥本海默起了这样一个颇具意味的名字。

[75]译注：荷兰米德尔堡的眼镜商，发明了望远镜。

[76]其中之一是计划中的激光干涉仪空间天线（Laser Interferometer SPace Antenna, LISA），LIGO的太空版，由多个彼此间隔百万千米的飞船组成，这些飞船扮演着组成LIGO的管子的角色。LIGO也有可能与VIRGO合作，VIRGO是法国意大利联合运行的引力波探测器，坐落于比萨城外。

[77]译注：由于技术上的原因，截至2008年年初，LHC并未启动，而是于2009年才开始运行。目前运行良好。

[78]译注：纽约扬基队的终结投手，以投快速球著称，被称为扬基守护神。

[79]译注：或者说差不多有3个香港那么大。

[80]译注：在国内的文献中，超电子、超夸克、超中微子也会译作标量电子、标量夸克和标量中微子，因为自旋为0的粒子是所谓的标量粒子。

[81]译注：每年十二月，诺贝尔奖都会在瑞典首都斯德哥尔摩颁发。

[82]译注：美国著名电视剧《星际迷航》中的飞船。

[83]译注：美国作家约瑟夫·米勒的名著《第二十二条军规》使“第二十二条军规”这个词进入英语，用以指无论怎么做都不行的限制性条款。

[84]超距传输要讨论的是将一个处于此处的物体传输到别的地方，因此，在本节探讨有关问题时我说的话常常会带来一种感觉——好像粒子可以有确定的位置似的。事实上，更准确的说法应该是“一个有很大概率处于这一位置的粒子”或者“有99%的概率处于这一位置的粒子”，当然，谈到一个粒子被传输到某一位置时也应该这么说。但是为了不这么罗里八唆，我就用了不太严格的语言。

[85]译注：DeLorean，汽车品牌。科幻电影《回到未来》中的时间机器就是用这个牌子的运动型轿车改装而成的。

[86]译注：联合海陆货运公司，United Van Lines，美国物流公司。作者在这里改用了其名称。

[87]对于粒子集合来说——与单个粒子不同——量子态也需要囊括这个集合中每个粒子之间彼此关系的有关信息。所以，要想准确的复制组成一台DeLorean的粒子的量子态，我们必须确保这些粒子之间的关系也被正确的复制；在这个过程中，唯一的变化只能是这些粒子的位置——从纽约变到了伦敦。

[88]脆弱的人体也是限制之一：在合理的时间内达到所需要的速度必须要有非常大的加速度，这样的加速度可能远远超过人体的承受能力。还要注意到，时间变慢使我们有可能——至少有理论上的可能性——到达太空中非常远的地方。要是一颗火箭离开地球后以99.99999999999999999%的光速朝着仙女座星系飞去，那么我们可能就得等600万年才能看到它的归来。但在这样的速度下，火箭上的时间将比地球上的时间慢得不可想象。对火箭上的宇航员来说，回到地球的时候，整个行程刚刚用掉8小时（我们姑且先把宇航员根本承受不了达到这样的速度而必须有的巨大加速度这件事放在一边，先假定他或她可以承受这样的加速度，并且安全返航）。

[89]当然，我实际上应该说1966年1月1日，不过没关系。

[90]译注：H.G.威尔斯（1866～1946），英国著名的科幻小说家。名著《时间机器》是他的第一部科幻小说；布朗博士，电影《回到未来》三部曲中的疯狂科学家，他把自己的DeLorean轿车改装成了时间机器。

[91]译注：这两个地名都是动画片《辛普森一家》中的内容。其中Kwik-E-Mart是一家连锁超市。

[92]译注：Kwik-E-Mart便利店的印度裔老板。

[93]译注：Eureka意为“明白了”“搞清楚了”。据说希腊著名学者阿基米德在洗澡的时候突然想清楚了测量不规则物体体积的办法，兴奋地跳出浴缸，高呼：“Eureka！”

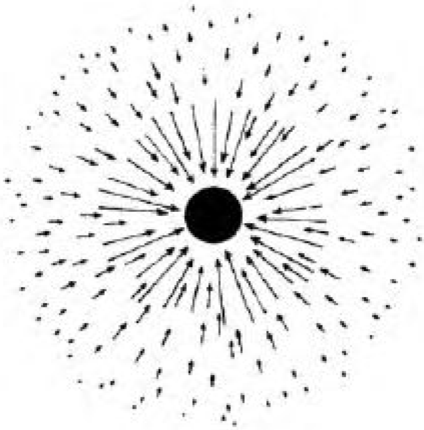
[94]要是对与圆环以及卡拉比丘流形有关的几何对偶性的细节有兴趣的话，请参见《宇宙的琴弦》第10章。

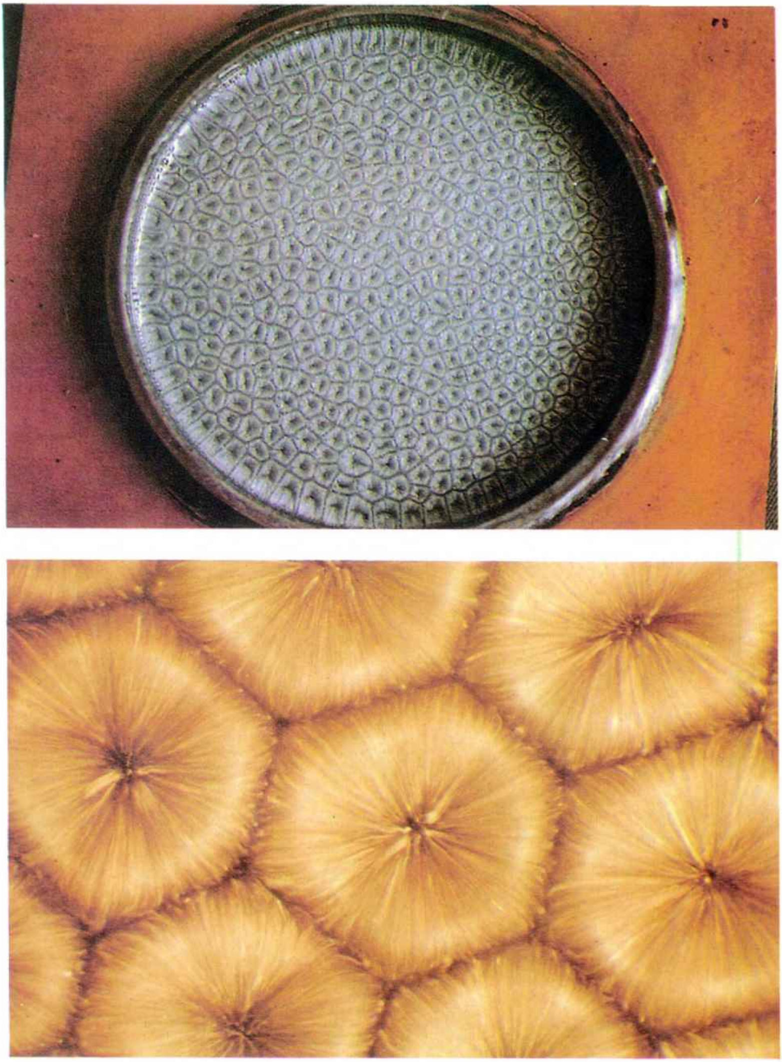
[95]译注：特百惠，美国知名家用塑料制品品牌。

[96]如果你不愿意再提柏拉图，那么就可以用膜世界原理所提供的全息图说法来理解整件事。在这个说法中，影子就是它本来的意思。假想我们生活在包裹于四维空间外面的3-膜上（就像三维的苹果外面的二维苹果皮）。全息原理说的就是我们的三维感知实际上是发生在我们的膜世界所包裹着的四维世界的物理的影子。

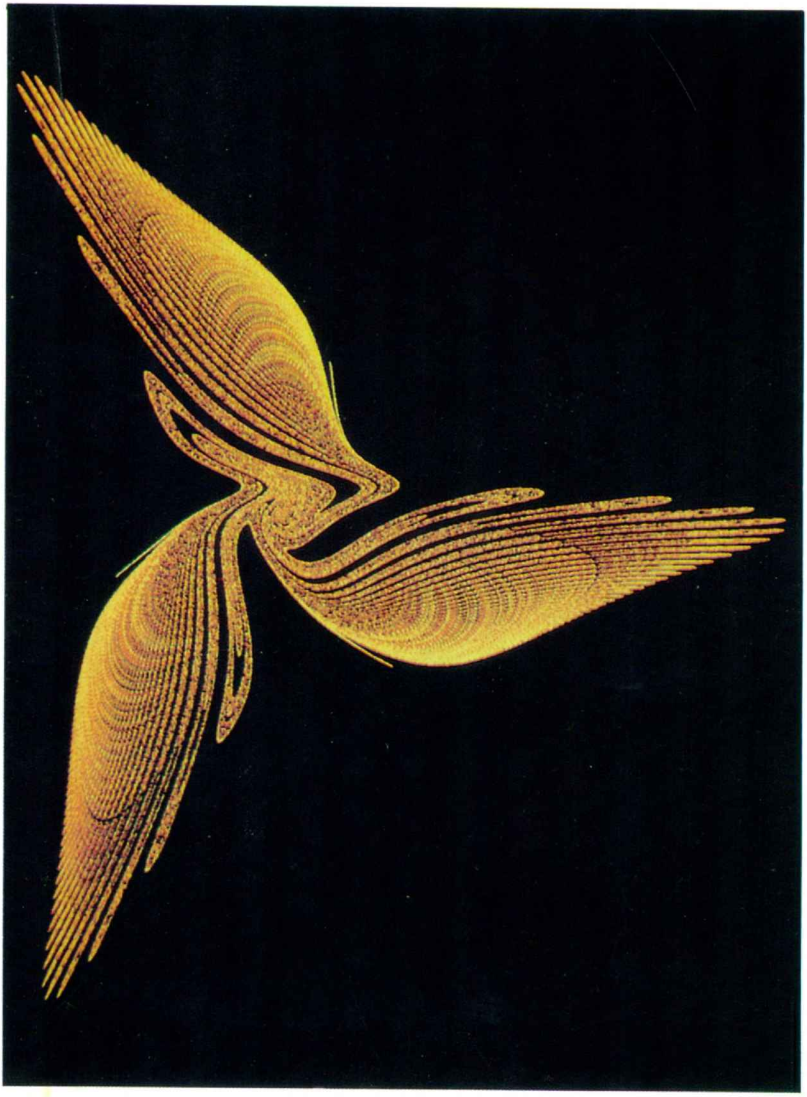
[97]译者注：这些书中，只有部分有中译本，国内出版的可能收录不全。

# 时间之箭

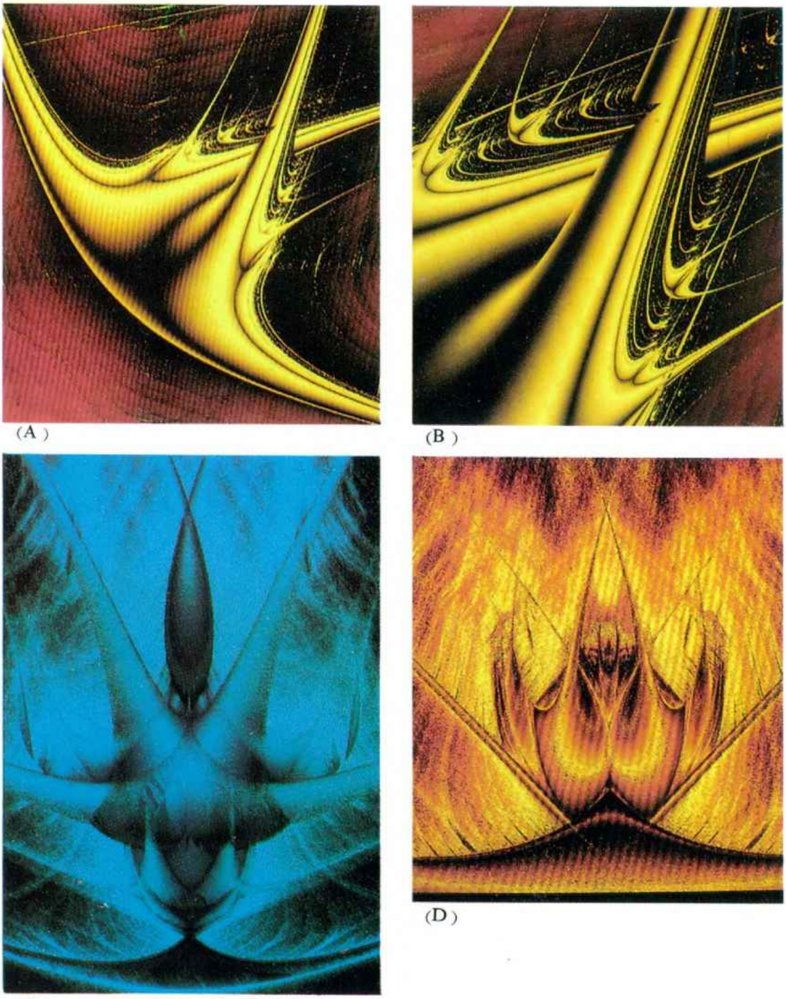




一层硅黄，一毫米厚，其中的六角形对流图案的全貌（上）与特写（下）



奇异吸引子一例

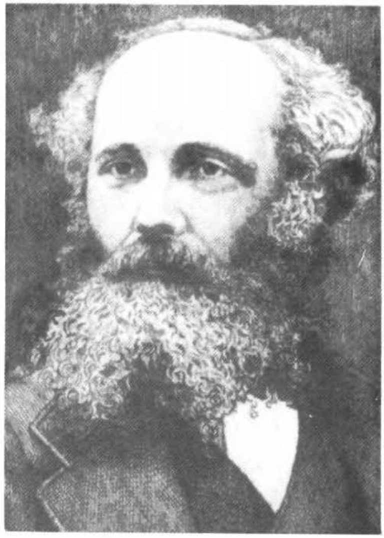


(A) (B) (C) (D)

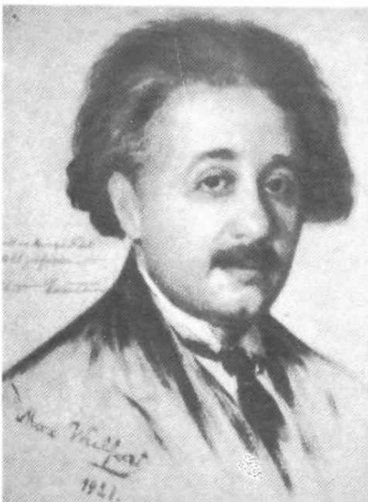
筹运方程式图示。一个群体总数动力学的简单模型的操作表演。A 和B：参数相同、放大率不同。C 和D：不同参数、同一放大率。淡区对应组织，深区对应混沌。



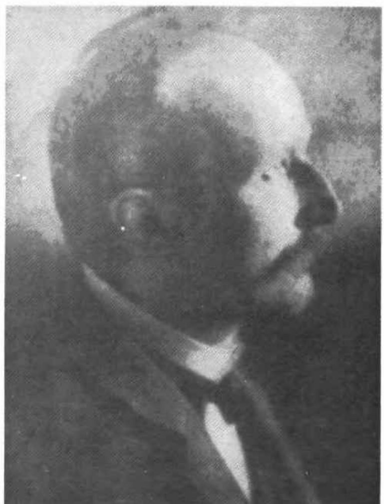
路德维格·玻耳兹曼



詹姆斯·克拉克·马克斯韦



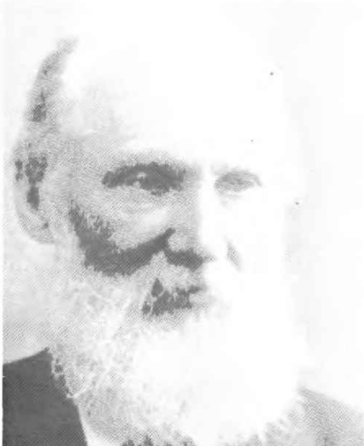

阿尔伯特·爱因斯坦



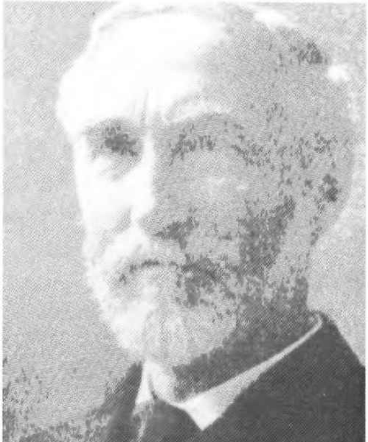
麦克斯·普朗克



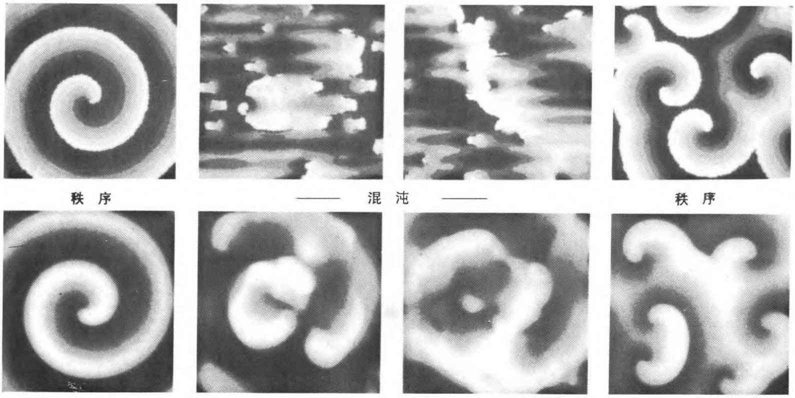
尼古拉斯·莱纳德·薩迪·卡诺



开尔文爵士·威廉·汤姆逊



J·威拉德·吉布斯



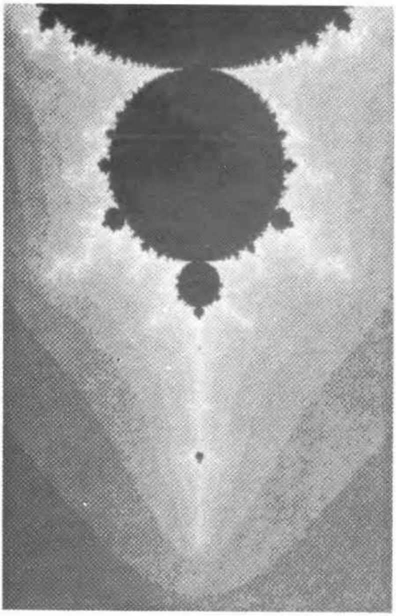

贝鲁索夫——札索廷斯基反应从秩序中滑进混沌，又滑回来。上行是电脑模拟，下行是反应本身。（下）一个分维体。不管放大多少，形状都是一样。

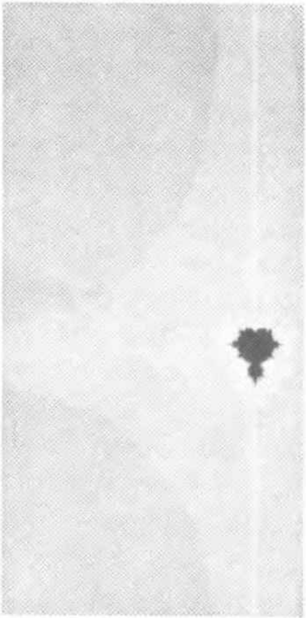

拉尔斯·昂萨格

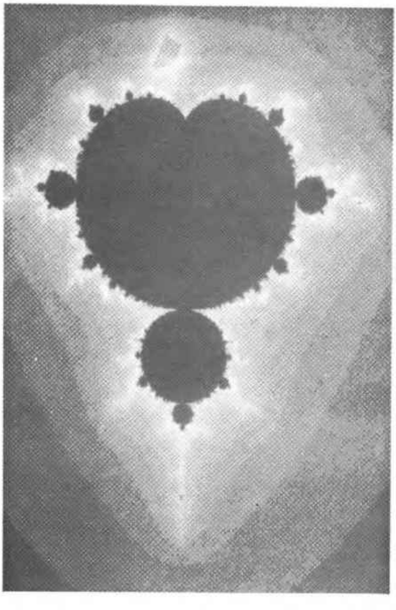

伊里亚·普里高金

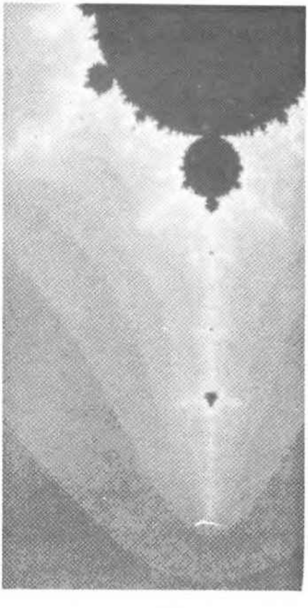

茹道尔夫·克劳修斯

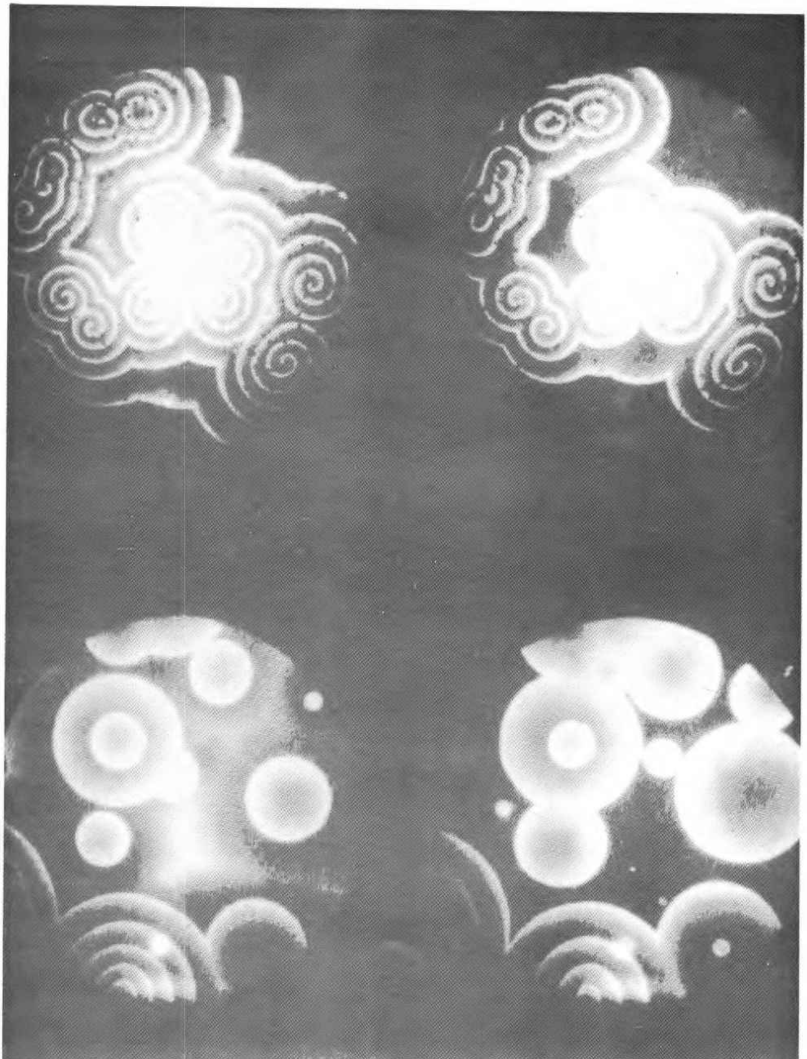

詹姆士·普莱斯特考·焦耳

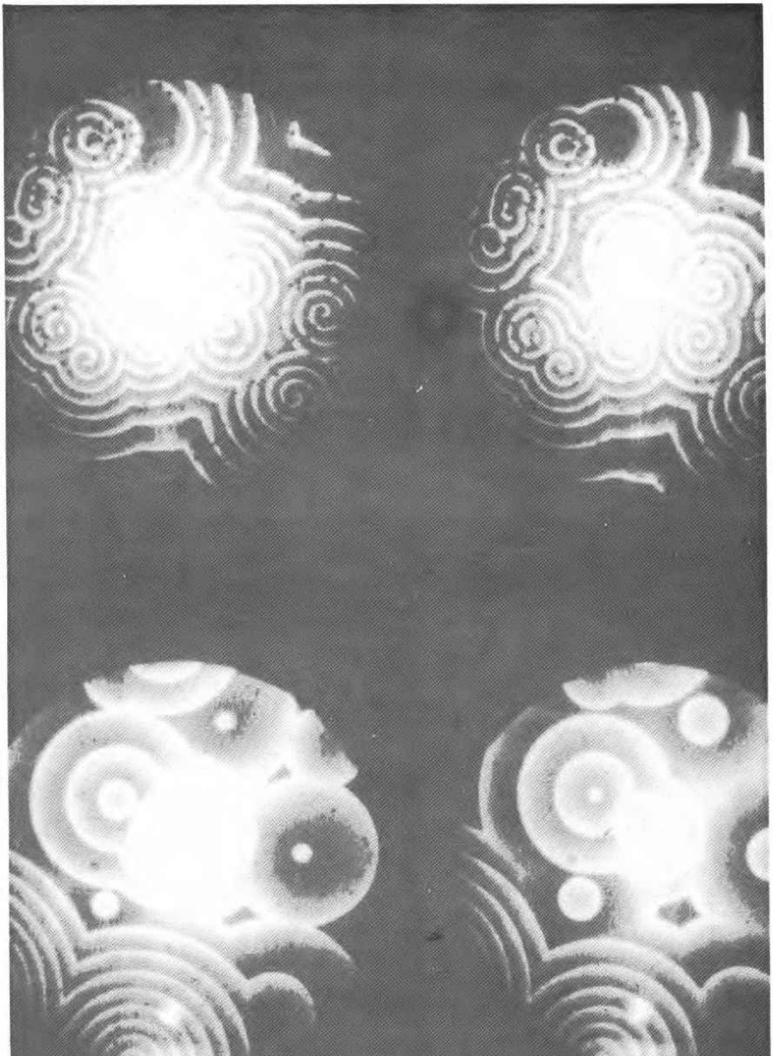




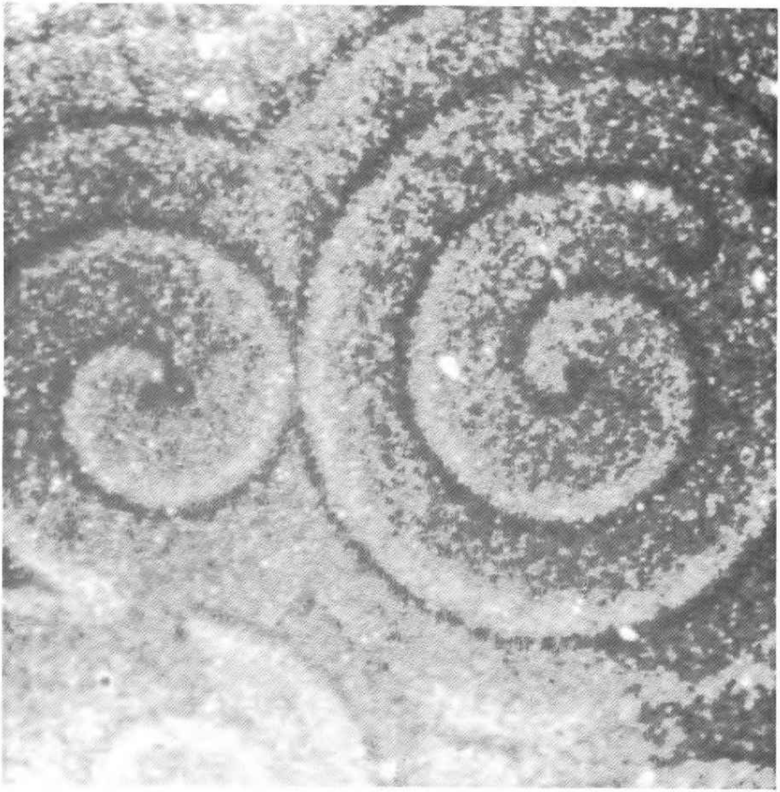








贝鲁索夫——札索廷斯基反应中形成的螺旋花样。



粘菌 Dictyostelium discoideum 聚合

# 总序

科学，特别是自然科学，最重要的目标之一，就是追寻科学本身的原动力，或曰追寻其第一推动。同时，科学的这种追求精神本身，又成为社会发展和人类进步的一种最基本的推动。

科学总是寻求发现和了解客观世界的新现象，研究和掌握新规律，总是在不懈地追求真理。科学是认真的、严谨的、实事求是的，同时，科学又是创造的。科学的最基本态度之一就是疑问，科学的最基本精神之一就是批判。

的确，科学活动，特别是自然科学活动，比较起其他的人类活动来，其最基本特征就是不断进步。哪怕在其他方面倒退的时候，科学却总是进步着，即使是缓慢而艰难的进步。这表明，自然科学活动中包含着人类的最进步因素。

正是在这个意义上，科学堪称为人类进步的“第一推动”。

科学教育，特别是自然科学的教育，是提高人们素质的重要因素，是现代教育的一个核心。科学教育不仅使人获得生活和工作所需的知识和技能，更重要的是使人获得科学思想、科学精神、科学态度以及科学方法的熏陶和培养，使人获得非生物本能的智慧，获得非与生俱来的灵魂。可以这样说，没有科学的“教育”，只是培养信仰，而不是教育。没有受过科学教育的人，只能称为受过训练，而非受过教育。

正是在这个意义上，科学堪称为使人进化为现代人的“第一推

动”。

近百年来，无数仁人智士意识到，强国富民再造中国离不开科学技术，他们为摆脱愚昧与无知作了艰苦卓绝的奋斗，中国的科学先贤们代代相传，不遗余力地为中国的进步献身于科学启蒙运动，以图完成国人的强国梦。然而应该说，这个目标远未达到。今日的中国需要新的科学启蒙，需要现代科学教育。只有全社会的人具备较高的科学素质，以科学的精神和思想，科学的态度和方法作为探讨和解决各类问题的共同基础和出发点，社会才能更好地向前发展和进步。因此，中国的进步离不开科学，是毋庸置疑的。

正是在这个意义上，似乎可以说，科学已被公认是中国进步所必不可少的推动。

然而，这并不意味着，科学的精神也同样地被公认和接受。虽然，科学已渗透到社会的各个领域和层面，科学的价值和地位也更高了。但是，毋庸讳言，在一定的范围内，或某些特定时候，人们只是承认“科学是有用的”，只停留在对科学所带来的后果的接受和承认，而不是对科学的原动力，科学的精神的接受和承认。此种现象的存在也是不能忽视的。

科学的精神之一，是它自身就是自身的“第一推动”。也就是说，科学活动在原则上是不隶属于服务于神学的，不隶属于服务于儒学的，科学活动在原则上也不隶属于服务于任何哲学的。科学是超越宗教差别的，超越民族差别的，超越党派差别的，超越文化的地域差别的，科学是普适的、独立的，它自身就是自身的主宰。

湖南科学技术出版社精选了一批关于科学思想和科学精神的世界名著，请有关学者译成中文出版，其目的就是为了传播科学的精神，科学的思想，特别是自然科学的精神和思想，从而起到倡导科学精神，推动科技发展，对全民进行新的科学启蒙和科学教育的作用，为

中国的进步作一点推动。丛书定名为《第一推动》，当然并非说其中每一册都是第一推动，但是可以肯定，蕴含在每一册中的科学的内容、观点、思想和精神，都会使你或多或少地更接近第一推动，或多或少地发现，自身如何成为自身的主宰。

《第一推动》丛书编委会

# 前言

1977年诺贝尔化学奖获得者

伊里亚·普里高津

我十分高兴在此为彼得·柯文尼、罗杰·海菲尔德的这本书写篇前言。

时间有箭头吗？这问题自从苏格拉底以来，一直迷魅着西方的哲学家、科学家和艺术家。然而，在20世纪末期的今日，我们问这问题，情况与以前不同。对一个物理学家来说，20世纪的科学史可以分为三个阶段。首先是两项思想方案——相对论和量子力学——所产生的突破。其次是一些出人意料之外的事物的发现，包括“基本”粒子的不稳定性、演化宇宙论，以及包括诸如化学钟、决定性混沌等的非平衡结构。最后——也就是现在，由于这些新的发展，我们必须对整个物理学作重新思考。

这里一个令人瞩目的特点是：所有这一切都强调着时间所扮演的角色。当然，在19世纪，人们都已经承认时间在生物学、社会科学等学科中的重要性。可是当时一般认为，物理描述的最基本层次是可以用决定性的、时间可逆的规律来表达，而时间箭头只相当于唯象层次的描述。这种立场在今日是很难站得住了。

现在我们知道，时间之箭在非平衡结构的形成之中，扮演着最重要的角色。近来的研究告诉我们，这些结构的演化可以在计算机上用按照动力学规律写的程序来模拟。因此很显然地，自我组织过程不会

是某些唯象观假设的结果，而是内禀于某类动力系统之中的属性。

熵的意义，我们现在更能体会了。按照热力学第二定律，熵这个量总是在增加的，因此它赋予时间一个箭头。熵基本上是高度不稳定系统所具有的一个性质。这种系统，将在此书第六章和第八章详加讨论。要研究的东西还很多，许多问题仍是悬案。因此不足为奇，我不一定同意这本书里的每一句话。但是对作者所提倡的一般立场，我是同意的，即时间之箭是某些重要种类的动力系统中一个精确的性质。

这些问题是非常重要的，因此我热烈欢迎此书的写成。这本书科学水平很高，而同时能被较多的读者接受。彼得·柯文尼本人对此领域作过重要的贡献，因此撰写此书，尤其胜任。罗杰·海菲尔德的文笔流畅，使此书精彩可读。

1989年10月在明尼苏达州圣彼得城的古斯塔乌斯·阿道尔夫斯学院举行的诺贝尔会议，专门讨论了一个充满挑战性的题目：“科学的终结。”会议组织人写道：“越来越有这样的感觉……科学已不再能被当做一种统一的、普遍的客观努力。”他们接着写道：“如果科学只搞‘超历史的’普适的定律，而不理会社会性的、有时间性的、局部的事物，那我们就无法谈及科学本身以外某些的真情实况，而科学仅仅是反映而已。”这句话把“超历史的”规律和有时间性的知识对立起来。科学的确是在重新发现时间，这在某种意义上标志着对科学的传统看法的终结。但这难道是说科学本身完结了吗？

的确，我上面已经提到，经典科学的研究方针是把全力集中在用决定性的、时间可逆的规律来描述世界。实际上，该计划从未完成过，这是因为，规律以外还需要事件，而事件在对自然的描述中引入时间之箭。屡次三番，经典科学的目的似乎就快完成了，但结果总是出了岔子。这种情况给予科学史几分戏剧性的紧张。例如，爱因斯坦的目的是把物理学表达为自然界的某种几何，可是广义相对论给现代宇宙论开路以后，遇到的却是所有事件中最惊人的事件：宇宙的诞

生。

“规律一事件”二重性是西方思想史中一直在进行的争论的中心，这争论从苏格拉底以前的臆想，经过量子力学和相对论一直进行到我们这时代。伴随规律的是连续的展开，是可理解性，是决定性的预言，而最后是时间本身的否定。事件却意味着某种程度的任意性、概率以及不可逆的演化。我们必须承认我们是居住在一个二重性的宇宙里面，这宇宙既牵涉到规律，也牵涉到事件，既有必然，也有或然。我们所知道的事件之中，显然最关键的是和我们宇宙的创生和生命的形成有关联的事件。

“我们有一天能克服热力学第二定律吗？”——阿西莫夫（Asimov）的科幻小说《最后的问题》中的世界文明不断地在问一台巨型计算机。计算机回答道：“资料不足。”亿万年过去了，星辰、星系都死了，而直接和时空联结的计算机仍在继续搜集资料。最后没有任何资料可以搜集了，不再“存在”任何事物了；可是计算机还是在那儿计算，在那儿找相关关系。最后它得到答案了。那时候要知道这个答案的人也都不存在了，可是计算机知道了如何克服热力学第二定律。“于是光明出现……”对阿西莫夫来说，生命之出现、宇宙的诞生都是“反熵”的、非自然的事件。

此书是新思维框架的一个极好的介绍。这个新思维框架将导致一套既包括规律又包括事件的新物理学，将使我们对我们身处的世界有更好的了解。

# 序

地点：的里亚斯特附近的都伊诺小镇

日期：1906年9月5日

路德维格·玻耳兹曼（Ludwig Boltzmann）那时在亚得里亚海边的一个村庄里度假。这假期的用意是让他从在维也纳的研究工作中散散心，帮他康复已有一段时期的病和排遣心情的忧闷。可是玻耳兹曼心绪还是很不安宁。

人的心理中有一个基本假设，那就是，时间一去不回头。对此假设，只存在一个科学证据。玻耳兹曼自从他20来岁时当了教授以来，多年来就是为了想了解这证据而苦斗。此项伟大的追求，他没有成功。他有关“熵”的工作（熵是衡量变化的一个物理量，它总是随时间增长的）非常精彩，可是没有得到确定的结果。时间方向之谜始终是科学的缺陷。而对玻耳兹曼来说，时间已经到了尽头了。

玻耳兹曼身材魁梧，一脸的大胡子，可是人不可貌相，其实他性格柔弱，容易受人伤害。他工作过度，疾病缠身。那年他62岁，双目差不多完全失明，剧烈的头痛使他坐卧不安。起伏的情绪曾一度把他带到绝望的边缘，使他在慕尼黑附近的一个疯人院里住过一段时期。一点不顺心的事也会使他大大伤心——例如今天他夫人为了要把他的外套拿去洗衣店干洗，因而坚持他晚一些时候再回维也纳。

玻耳兹曼夫人拿走了外套，和她女儿一同去西司提亚纳海湾游泳去了。就是在那个时候，她丈夫做了天下最不可逆转的事。他把一根

短绳子系在窗框的横木上，围着自己的脖子打了一个死结。然后，就在那间租住的屋子里，他自杀了。他女儿艾勒萨回来，看见父亲在那儿上了吊。

玻耳兹曼的自杀，再次生动地表现了时间是如何在捉弄想揭示时间秘密的人。他的丧生深深地震撼了人们的心灵。他在莱比锡的一个学生嘉菲（George Jaffé）写道：“玻耳兹曼的死是科学史中的悲剧之一，就像拉瓦锡（Lavoisier）上断头台，迈耶（R.J.Mayer）进疯人院，皮耶尔·居里（Pierre Cu-rie）惨死在货车轮下一样。尤其可悲的是：这事就发生在他的思想将取得最后胜利的前夕。”

玻耳兹曼的思想牵涉到原子是否存在。有些评论家把玻耳兹曼看做一个智力的“三十年战争”的受难者，这战争是和不肯接受原子论的人打的。他的对手包括一大批19世纪思想界中的知名人士，其中有法国的督黑姆（Pierre Duhem）、孔德（Auguste Comte）和庞加莱（Henri Poincaré），德国的奥斯特瓦尔德（Wilhelm Ostwald）、荷耳姆（Goerg Helm），以及其他在美国和英国的诸如兰金（William Rankine）、史大罗（John Stallo）等人。他最大的敌手是他的同胞马赫（Ernst Mach），他们之间的论战使玻耳兹曼在知识界中形势孤立。玻耳兹曼有一次曾向一个同事表白，说懂他理论的人，实在是一个也没有。

玻耳兹曼对原子和分子的看法，终于占到上风。可是他希望能进一步解释时间的方向。对于自然界的这个特点，玻耳兹曼比其他任何科学家都下了更大的工夫。他这方面的雄心，终被狂躁抑郁症所击败，使他自杀。下面我们将要看到，玻耳兹曼在原子分子的概念和时间的方向之间，成功地建立了一个决定性的关联。可是他的伟大梦想，在他生前始终没有实现。

在努力想用原子分子语言来表达时间箭头和我们这世界其他特性的人中，玻耳兹曼还不是最后一个死于悲惨结果的人。加州工学院的

古德斯坦（David　Goodstein）在他的《物质的状态》一书的开头写道：“一生中大部分时间花在研究统计力学的路德维格·玻耳兹曼，1906年自杀身亡。艾伦菲斯特（Paul　Ehrenfest）继续了这项工作，但也以类似的方式而死。现在轮到我们了……也许研究此项课题，谨慎为妙。”

# 第一章　时间的形象

毁败教我这样想来想去

时间要来把我所爱带走

这念头好像死亡，不得不

为所害怕的丧失而哭

——莎士比亚

《十四行诗》

时间是给人神秘感最大的来源之一。它深奥难测的性质，是有史以来人们日夜捉摸的对象。历代的诗人、作家、哲学家都被时间迷惑过。可是，近代的科学家们却没有这样。现代科学，尤其是物理学，即使没有完全取消，也总在想降低时间在事物中的作用。因此有人称时间为被忘却的维度。

我们都知道时间一去不返，觉得它的流逝好像支配着我们的存在，过去已不可改变，未来是一片空白。我们有时巴不得能扳回时针，能挽回过失，能重享美好的时光。可惜，常理不允许我们这么做。我们知道，时间是不等人的，时间不会倒流。

真不会吗？奇怪的是，许多科学理论并不支持我们一般对时间的看法；在这些理论中，时间的方向无关紧要。如果时间倒走，现代科学的几座大厦——牛顿力学、爱因斯坦的相对论、海森堡和薛定谔的量子力学，也都同样站得住脚。对这些理论来说，记录在影片上的事

件，不管影片顺放还是倒放，看上去都行得通。单向的时间，反而像是我们脑中产生的幻觉。研究这个问题的科学家们，带着几分嘲笑的口气，把我们日常时间流逝的感觉，称为“心理时间”或者“主观时间”。

宇宙会不会有这样一个地方，那里时间的方向跟我们所熟悉的方向相反，那里的人们从坟墓里升出，皱纹从脸上消失，然后回进母胎？在那个世界里，香气神秘地凝结成香水，钻入瓶中；池塘里的水波向中心汇聚，弹出石头；屋里的空气自发地把各种成分分解出来；破瘪的橡皮膜自动膨胀，密封成气球；光从观测者的眼睛里射出来，然后被星球吸收。可能的事或许还不止这些。按照这个想法，地球上的时间也会开倒车，我们也都会被过去所吞没。

那样就跟所有时间总朝一个方向走的大量事实完全矛盾了。让我们比较一下时间和空间。空间包围着我们四周，而时间总是一点一点地体验到。左右之间的差别，根本比不上过去和未来之间的差别。在空间中我们可以朝四面八方走来走去，而我们一切行动只能对将来起作用，不能影响过去。我们只有回忆，但除非是千里眼，不能预知未来。物质一般总是逐渐地腐烂下去，而不会自发地聚合。这样看来，特殊的方向，空间没有，时间有。时间行走像一支箭。“时间之箭”——这意味深长的词，是英国天体物理学家爱丁顿（Arthur Eddington）在1927年首先提出的。

本书探讨时间在当前科学理论中扮演的各种角色及其后果，并且指出我们的确能找到一个对时间的统一看法，这个看法和我们直接经验的时间一致、不矛盾。时间之箭甚至会启发我们：为了描述自然界，有必要建立一个比目前更深刻、更基本的理论框架。

## 文学中的时间

平常一般对时间的看法，在一些文学名著中得到极生动的表达。单向的时间使我们觉得万事无常；这种感情，再好不过地流露在普鲁斯特（Marcel Proust）自传小说的书名“重找失去的时光”里。这些作家脑子里转来转去的，就是人生短促有限，光阴一去不返。时间不由自主地向前走，每个时刻我们都得抢，都得尽情玩味。昙花一现，生命的神秘更加神奇；朝生暮死，更使我们觉得时间的不可逆。象征时间的老人，代表死亡的骷髅收割者，同样带有一把镰刀、一瓶沙漏，这不是偶然的：时间到了，谁也逃不过那把镰刀。

多少诗、多少文章，是描述时间的流逝！古波斯哲学家兼诗人奥玛·哈央姆（Omar Khayyam，死于1123年）的冥思，在菲兹哲罗（Edward Fitzgerald）的意译之中永存不朽：

不停前移的手指写字

字一写完就向前去

不管你多虔诚多聪明

它不会回来改半句

你眼睛里所有的眼泪

洗不掉它的一个字

人生的悲哀归根结底来自时间的不可逆转。不言而喻，最后胜利属于死亡。凡是活着的都要死，这事实就是时间流逝的铁证；这里就和科学开始挂钩。要想了解我们周围的世界，这一点非解决不可。就如爱丁顿所说：“在属于内心和外界的两种经验之间搭任何桥梁，时间都占着最关键的地位。”

## 文化时间

时间有向的概念，并不是一直都有的。潮水、冬夏二至、季节、星辰的循环往来，这些现象使许多原始社会把时间看做一种基本上不断循环的有机节奏。他们想，既然时间跟天体的循环运转分不开，时间本身也应该是循环的。白天跟随黑夜，新月代替旧月，冬天过了是夏天，为什么历史就不这样？中美洲的玛雅人相信历史每260年重复一次，这个周期他们叫拉马特，是他们日历的基本单元。他们认为灾难也有周期：1698年，西班牙人入侵登陆，伊嚓部落闻风而逃，因为他们相信周期满了，灾难来到。这点他们并没有搞错，但并不是什么预言，连巧合都算不上。原因是入侵80年前，西班牙人从传教士那边得知玛雅人相信时间有周期，所以侵略者本来就预料到对方的反应。

时间的循环模式是希腊各宇宙学派的一个共同点。亚里士多德在他的《物理学》中说："凡是具有天然运动和生死的，都有一个循环。这是因为任何事物都是由时间辨别，都好像根据一个周期开始和结束；因此甚至时间本身也被认为是个循环。"斯多葛（Stoics）学派的人相信，每当行星回到它们初始相对位置时，宇宙就重新开始。公元4世纪的尼梅修斯（Nemesius）主教说过："苏格拉底也好，柏拉图也好，人人都会复生，都会再见到同样的朋友，再和同样的熟人来往。他们将再有同样的经验，从事同样的活动。每个城市、每个村庄、每块田地，都要恢复原样。而且这种复原不仅是一次，而是二次三次，直到永远。"好像所有历史的事件都装在一个大轮子上一样，循环不已。这不断回返的观念重新出现在现代数学里面，叫"庞加莱循环"。庞加莱（Henri Poincaré）是世界伟大数学家之一，活跃于19世纪末、20世纪初。

时间之箭引起我们内心的恐惧，因为它意味着不稳定和变迁。它所指向的是世界的末日，而不是世界的重新再生。罗马尼亚人类学者、宗教史学者埃里阿德（Mireea Eliade）在他有关时间之箭和时间循环、名为《永恒回返的神话》的书里，认为世上从有人类以来，多半的人都觉得循环时间更令人安慰，而将它紧抱不放，死心塌地地承

认再生和更新；这样，过去也是将来，没有真正的“历史”可言。请注意他写的：“远古人的生命……虽然发生在时间里面，并不记录时间的不可逆性；换句话说，对时间意识中最明确的特征，它反而置之不理。”

是犹太基督教传统把“线性”（不可逆）的时间，一下子直截了当地建立在西方文化里面。埃里阿德写道：“这种‘无尽循环’的老调，基督教企图一下子将它超越。”由于基督教相信耶稣的生、死和他的上十字架受难，都是唯一的事件，都是不会重复的，西方文化终于把时间看成是穿越在过去和未来之间的一条线。基督教出现以前，只有犹太人和信仰拜火教的波斯人认同这种前进式的时间。

不可逆时间深刻地影响了西方思想。对“进步”和地质学所谓的“深时”——指人类进化只是新上地球舞台不久的一出戏的那项惊人发现，不可逆时间给我们做了心理准备。它为达尔文的进化论开辟了道路，从而把我们和原始生物在时间上连接起来。总之，线性时间概念的出现，和因之而起的思想改变，为现代科学以及其改善地球上生命的保证打下了基础。

文化时间的循环模式和线性模式，在生物时间中可以找到对应。细胞的分裂，以及体内各种不同节奏——从高频的神经脉冲到悠闲的细胞更新——所组成的交响乐，都牵涉循环式时间；而不可逆的时间则体现于从生到死的老化过程之中。日常用的钟表也具有这两个不同的时间面貌。一方面，不停的钟摆或晶体振荡累积成一般所谓的“时间”，在地球上这时间就表示为12小时或24小时的周期。另一方面，各种耗散现象，诸如电池的干涸、发条的松弛、钟锤的下降，都告诉我们时间是一去不回头的。

## 哲学中的时间

时间是哲学家不断思索研究的一项课题。数学家惠特罗（Gerald Whitrow）在他的名著《时间的自然哲学》里面，强调阿基米德和亚里士多德是对时间两个极端看法的代表人：亚里士多德认为时间是内禀的，对宇宙说来是基本的，而阿基米德的看法就完全相反。2000年来，这个争论，以不同的形式一直在进行。

柏拉图在他讲宇宙学的著作《狄玛尤斯》（Timaeos）里说，太初混沌，神工强加以形序，时间乃生。《狄玛尤斯》一开头就讲“实存”（being）和“将然”（becoming）的区别，这两个概念，以各种形式重新出现于近代科学理论之中。对柏拉图，“实存”的世界是真正的世界，“此世界永恒不变，由智慧借助论证而得知”；而“将然”的世界（时间的领域），则是“意见与非理智感觉之客体，既生又灭，从未完全真实过”。他把“将然”比作旅途，把“实存”比作终点，说只有后者才是真实的。这样，具有时间的物理世界只有次等的真实性。这种区别，支配着柏拉图哲学的全部。

拥有这种看法的人，在柏拉图以前是帕尔米尼笛斯（Parmenides）。他相信实际是既不可分，也没有时间的。他的学生，意大利南部埃里亚的芝诺（Zeno），为了整个推翻我们对时间的观念，创造了一批著名的佯谬来捉弄我们。其中最有名的一个叫“勇士和乌龟”，企图用来证明：“如果时间可以无穷地分而再分，运动将是不可能。”设想勇士在追乌龟，当勇士达到当初乌龟在的地方，乌龟又走了一（小）段路；当勇士把那段路走完后，乌龟又走了一段，这样下去，永无止境。

究竟芝诺这个佯谬和他别的佯谬意义多大，众论不一。自从芝诺在24个世纪以前把它们提出以来，至今已有大批文献问世，有的说是无稽之谈，有的却认为非常高深。惠特罗仔细分析的结论是：要解开这些佯谬，只有两条路。一条是否定“将然”这个概念，于是时间便有真正属于空间的性质；另一条路是不承认时间像空间那样可以分而再分。

就像红颜色给不同的人有不同的主观印象，而仍然是视觉中不可少的一部分一样，哲学家康德同样认为，时间固然是我们经验中不可缺少的一个成分，它其实是没有客观意义的：“时间并不是什么客观的东西，它既不是实体，也不是偶发，也不是关系；它是因为人类心灵的本性而必然产生的主观条件。”康德的“主观主义”看法，和有些科学家对今日科学中时间的解释十分相似。一个简单的、显而易见的、被历代的唯心主义者们——帕尔米尼笛斯、柏拉图、斯宾诺扎、黑格尔、布拉德里（Bradley）、麦克太戈特（McTaggart）等——所乐于采取的办法，是说时间充满矛盾，所以不会是真实的。对这种形而上学的遁词，逻辑学家克琉（M.Cleugh）老实不客气地说道：“说时间由于自相矛盾便只是表面现象，非但没有解决问题，连答案都说不上。”

玻尔兹曼（Ludwig Boltzmann）给形而上学起了个诨名，叫它“人脑中的偏头疼”。他说：“最平常的东西一到哲学便成为不可解决的难题。哲学以无上的技巧建造了空间和时间的概念，然后又发现这样的空间里不可能有物体，这样的时间里不可能有过程发生……把这个叫做逻辑，我看就像一个想去山区旅行的人，穿着长而累赘的衣服，在平地上没走三步就被绊倒了一样。这种逻辑来源于对所谓的思想规律的盲目信任。”玻尔兹曼尖刻地批判了包括黑格尔、叔本华、康德的一群哲学家，他说：“为了追究到底，我拜读了黑格尔的著作，可是那里我看到的是滔滔不绝、不清不楚而又毫无思想的一番空话！从他那里总算我倒霉，我又去找叔本华……就是在康德那里，也有不少话叫我莫名其妙，以致令我怀疑，像他这样脑筋灵活的人，是否在跟读者开玩笑或是在存心欺骗读者。”

## 时间：牛顿和爱因斯坦

要是哲学里的时间使人失望，科学里的又如何？17世纪中叶，惠更斯（Christiaan Huygens）成功发明第一部摆钟，随后“计时”的精

度不断提高，逐渐令人觉得自然界是机械性的，是可预言的。时钟技术的发展把时间从人类的事件中解开，使我们更相信独立的科学世界。17～18世纪产生的“经典”科学所描绘的宇宙里面，自由意旨和偶发事件都是多余的；那个宇宙从各种观点来看，都无异于一台机器。

真正“科学时间”的诞生，我们可以上溯到发现物体运动数学表达式的牛顿。他的成就的确令人惊叹。他的运动公式，从苹果到月亮都能适用；地球上的动力学，天空上的动力学，被他融为一体。如此有力的表达式仅仅牵涉很少几个假设，给人十分优美的感觉，因此人们很快地接受了牛顿的思想。于是，牛顿成为现代科学的奠基者。

牛顿无疑受到数学家巴罗（Isaac　Barrow）的影响。1669年巴罗从剑桥著名的卢卡逊教授席退休时，设法保证该席由牛顿继任。巴罗曾说过，“既然数学家经常讲时间，他们对该词的意义应有明确的观念，否则他们不过是江湖术士罢了”。然而，尽管牛顿的科学成就如此辉煌，他方程式里的时间却只是一个未经定义的原始量。就像牛顿的空间一样，牛顿的时间也是绝对的。这就是说，任何事件，都在空间里有个一定的位置，都发生在时间里某个特定的时刻。格林尼治天文台也好，远处旋涡星系的一个角落也好，每个地方都被现在这个同一时刻所连接。牛顿在《自然哲学之数学原理》（以下简称《原理》）中说：“绝对的、真实的、数学的时间，由于它自身的本性……与任何外界事物无关地、均匀地流逝。”

牛顿力学具有极高的预测能力，它使每个时刻都有能力提供宇宙过去未来所有可能的信息。我们只要把宇宙某个时刻所有星球的位置、速度，放入一个解牛顿方程的巨大计算机就行了。冻结在那个时刻的，是整个的过去和未来：计算机能算出别的任何时刻星球的位置和速度。但是牛顿方程式不会做的是，它不能断定时间的哪个方向是我们宇宙的过去，哪个是未来。牛顿方程式从时间里把方向抽走了，没有为时间的不断前进性腾出任何地方。他这种对称式的时间，可以用拍摄行星运动的影片来阐明，例如用1977年发射的探测外太阳系的

旅行者二号飞船拍的影片。牛顿首次对这种运动定下了数学规律；然而，顺放也好，倒放也好，影片总是符合他的天体力学定律的。过去和未来都是预先注定的世界，对这种决定性世界的信仰，在物理学发展上起过极大的作用。它的影响可以从爱因斯坦接到他终生好友贝索（Michelangelo Besso）死讯时说的一句话里看出来。在他1955年3月21日的信里，爱因斯坦用他“物理定律没有时间性”这个坚强的信仰，试着给贝索的家人少许安慰。死并非终点，他写道：“对我们这些坚信物理学的人来说，过去、现在和未来之间的区别，尽管老缠着我们，不过是一个幻觉而已……”或许写这封信时，爱因斯坦也有意安慰自己，因为他又加了这么一句：“贝索向这个奇怪的世界告别，只比我稍早一点。”一个月后，爱因斯坦就去世了。

现在我们知道，牛顿的运动理论，在有些场合之下不适用：当物体的速度接近光速时，它不适用；在大质量、高引力场，包括黑洞的情况下，它也不适用；它也不适用于牵涉原子和亚原子粒子的极小尺度上。可是主管这些场合的20世纪两大革命，即爱因斯坦的相对论和量子力学，也同样建立在时间无向的概念上面。因此，要想在历史文学的不可逆时间和牛顿定律的对称时间之间搭桥梁，这两套理论同样是束手无策。

这并不是说它们没有对于时间提出新的、引人入胜的观念。爱因斯坦的相对论砸碎了牛顿绝对时间的通常观念——宇宙中任何事件都发生在空间的某一点、时间的某个时刻，而那时刻是到处一样的。爱因斯坦认为存在是四维的，是在合并三维空间和一维时间的四维时空中的存在，而不是一个三维存在外加它在时间上的演化。我们的时间感会被疾病或者药物捣乱。而爱因斯坦的相对论说，时间对不同运动状况的观察者是不同的：对某个观察者来说，一台钟如果钟本身移动得越快，它的时针、分针、秒针就走得越慢。相对论兴起以后，经过逻辑学家戈岱尔（Kurt Godel）的研究，连时间旅行的可能性，也登上了科学的大雅之堂。

尽管如此，爱因斯坦惊人的相对论对时间的单向性还是无话可说。就跟牛顿方程式一样，如果我们知道一个系统——例如围绕一个黑洞的一颗星，或者整个宇宙——在某个时刻的所有细节的话，那么爱因斯坦方程式的结构便能使我们知道它的整个过去和未来。可是，究竟哪个方向是过去，哪个是将来，它并未留下任何线索。此外，与它有关的数学中出现令人为难的“奇点”，在那里，空间、时间和物质无法描述，因而引起对它基本性的怀疑。最有名的奇点就是所谓的“大爆炸”——那个普遍被认为产生宇宙的超高密度火球。在这个奇点，巨大的能量都集中到一点，理论中的可测量都变为无穷大，从而变为无意义。正如宇宙学家西亚玛（Dennis Sciama）所说：“广义相对论蕴含着自戕的种子。”

## 量子时间

在我们给时间的方向寻找科学依据时，统治原子分子世界的量子论看上去似乎较有希望。量子论很成功地，同时也很令人不解地，描述了原子分子的各种奇行怪迹。它能解释激光、核反应器中的亚原子、计算机里的电子和许多其他东西的行为。我们从组成这个世界的原子分子大聚合的量子描述里，也许可以给我们有切肤感受的时间之箭，找到一个量子描述。这个想法来自一个很辉煌的传统。自从希腊文化的黄金时代以来，用组成世界的原子分子来描述世界——这部名为“归化原子法”的哲学遗产，在科学思想的发展中，占有至高无上的地位。

本书第四章将提到有两个棘手的问题：一是名为“长寿K粒子”的奇案，二是量子论中测量本身就是一个难解之谜。从这两点我们似乎能瞥见一个时间的量子箭头。可是，量子论的中心思想是跟随其他“基本”理论的，是不分时间的两个方向的。就和爱因斯坦的相对论一样，量子论也有严重的内症：用它处理实际问题时，譬如原子如何吸光、如何发光，它就会爆出许多令人头疼的无穷大。虽然物理学家

已理出一套能躲开这些无穷大的诀窍，我们也难免觉得内中大有蹊跷。

量子力学和爱因斯坦的相对论两虎不能容于一山。有些科学家，像牛津大学的彭罗斯（Roger Penrose），相信这两门理论如果真正统一了，就会出现一个量子引力理论，或者什么新的理论，其中时间之箭是明文规定的。这一步看上去一时还不能实现，并且很可能仍有不足之处。这是因为目前有一个危险，就是把我们的科学世界观死死地建造在具有时间可逆性的，原子、分子、粒子和场的微观层次上，而这个层次毕竟不是直接观测到的。正如诺贝尔奖获得者普里高金（Ilya Pri-gogine）所说："尽管物理大师跟我说了，我总还是不明白，怎么能从可逆性里得出我们宇宙、文化和生命的演化形式。"

## 时间与热力学

另一种描述牵涉科学家所谓的宏观层次。这个层次的现象是我们看得到、尝得到、碰得到和感觉得到的。热力学也就是探讨这个层次。19世纪，这门学问随着蒸汽动力的出现而问世，布莱克（Black）、卡诺（Carnot）、克劳修斯（Clausius）、玻尔兹曼、吉布斯（Gibbs）等人将它发展；它当时关注的是热机的功能。在一个形式的理论框架之下，热力学列出热和功之间的关系，详细说明了热如何转化为别种能量，如何跟别种能量交换。

我们对时间流逝的感觉，一面固然被经典力学、相对论、量子力学搞乱，另一面却从热力学中得到支持。就像我们觉得到时间是有方向的，热力学第二定律也说热只能从较热的物体流到较冷的物体，说雪人会融化，雕像会粉碎。第二定律和我们的时间感之间的关联可以用瓷器店和公牛的影片来说明。如果时间是朝正的方向走，影片就一定会显示出：完好的瓷器到处乱飞，碗碟被牛蹄踩碎。但如果我们看到公牛先倒走进被破坏得烂糟糟的店里，等到最后一个茶杯都好好地

飞回架上以后又倒走出来，那我们就知道影片放倒了。第二定律规定这种事不会发生，永动机不可能存在，证明任何过程中，能量都要转化成热而被消耗掉；这里，掉下来的碗碟的能变为再也不能复原的热和声。这种不能倒过来的损失是跟我们时间流逝的感觉连在一起的：从第二定律里面，我们发现一个叫做“熵”的量，它度量一个系统可变的能力，它跟时间有密切关系。熵的增大是时间方向的指路标。爱丁顿为了强调第二定律的至尊地位，发出了这样的警告：“……如果你的理论违背第二定律，那你就没有希望了，你的理论只有丢尽脸、垮台。”

上面我们看到，在牛顿力学里，过去、现在、未来的任何时刻都是一样的。因此，力学没有时间性，“演化”没有太深的意义。热力学就不同了，这里，熵把每个时刻加以区别，宇宙是真正在演化的。

就像古代的人，谈到不可逆时间就害怕，而一些哲学家们说时间是幻觉，草草地将它放在一边，很多科学家也很想把第二定律的含义简单埋葬了事。他们说，不可逆时间是和我们的头脑如何理解时间有关，而不是客观流逝。因为可逆时间的理论如此成功，所以这些科学家们想尽方法说第二定律中的时间之箭只不过是种幻觉。可是下面我们将会看到，这个说法如果对，那差不多所有的东西，连生命本身的节奏和过程，都变成我们个人的局限和近似的结果；这是因为解释有机过程的数学工具里面，时间之箭是基本的一部分。

一个常取的办法是，不讲别的，只问力学定律在实际上是怎样应用的。应用一个物理定律时，除掉这个定律以外，我们还得输入一些数字，代表初始条件，或者边界条件，例如宇宙中所有粒子的位置和速度。许多物理学家认为，热力学的时间之箭不知怎么搞的只是跟这些初始条件有关，而不是物理定律本身的一部分。

按照这个想法，如要了解时间之箭，应该考虑所有初始条件的老祖宗——宇宙的诞生。他们说，宇宙最初很小、很密，处于高度组

织、低熵状态，于是时间的流逝就一定和熵的增大、混乱的增大对应，而宇宙一直在膨胀，能量不断消耗成一片废热——所谓“宇宙热寂”的过程。但如果宇宙有朝一日开始缩小，这想法不是就不行了吗？那时候，熵不是就要开始降低了吗？时间不是就要开倒车了吗？下面我们会看到，这种牵涉边界条件的论调，说好一点，是主观性的，说坏一点，就是与题风马牛不相及。仅用初始条件显然解释不了时间之箭。另外有些人，其中包括数学物理学家霍金（Stephen Hawking），想把初始条件解释成宇宙论本身的结果。他的办法是把现今纯理论的对称时间的宇宙学理论推到尽头，然后说边界条件就是没有边界条件。这样一来，就不会出现不守规矩的奇点，空间时间也不会有“边缘”，宇宙将是独立自足的，犹如一个球面。

对宇宙学中的这类论调，南安普顿大学的兰兹堡（Peter Landsberg）不以为然，他写道：“诸如牛奶在咖啡中扩散这种‘日常’现象，反而要用本身比这些现象更有问题的、讲述宇宙的理论来‘解释’，这做法也未免太奇怪。一般人总认为一套东西应该用比它更肯定的另一套来解释；正常不可逆的现象用宇宙膨胀来解释，不属于这个办法。”

与其钻这个牛角尖，不如直接回到第二定律。对支持可逆（“不真实的”）时间的人来说，热力学最大的缺点是它只牵涉世界的皮毛，它不像相对论和牛顿力学那样能跟“基层的”、看不见的微观世界打交道。要对付这种批评，我们设想冰如何在可乐中融化的情况，办法是使可乐分子的动力学性质，以及诸如熵、体积、温度这类描述这杯可乐的宏观量与第二定律相一致。在这方面，玻尔兹曼作了出色的贡献，尽管许多他的同辈不相信他从原子分子的行为中，重新发现了时间之箭。玻尔兹曼震惊了当时的物理学家，他把熵和概率连在一起，成为世上第一个给一项基本物理定律一个统计性解释的人。他的开创性工作没有白做。例如，现在我们就可以用它来估计水分子在室温下，保持冰状态的概率：冰是个低熵状态，概率远小于水存在于液

态的概率——液态是随机性更大的高熵状态。表示这个关系的严密数学公式，爱因斯坦称它为“玻尔兹曼原则”；它已成为目前物理研究者广泛应用的一个工具。维也纳中区公墓里玻尔兹曼的坟上，就刻了这个公式，作为装饰纪念。虽然它没有回答我们的问题，但它为我们确立了研究方向。

乍看上去，第二定律和19世纪震惊全世界的另一发现——达尔文的进化论相冲突。经典力学把宇宙看成一个不折不扣的机器，热力学似乎说这机器步步走向绝对的混乱。可是达尔文证明的是，简单的生物逐渐演化成复杂的生物，生命随时间是越来越有组织，而不是越来越乱。天上飞的，水里游的，地上爬的，如此丰富多彩，它们的演化似乎跟主张世界越来越乱的理论不能相容。事实上，这里并不存在矛盾。这是因为热力学第二定律里面藏着一套妙计，它能使创造性的演化发生，而不仅只是纯破坏性的演化。早在1878年，玻尔兹曼可能已经看到了端倪，不过正式发展要等到近年来对第二定律的重新估价以后。新估价证明了第二定律并不意味着单调退化到无序，证明了宇宙反过来可以利用热力学来创造，来进化，来发展。这个新估价赋予第二定律的时间之箭新的微妙的意义，甚至更高的可信性。

## 创造性时间

以普里高金为首、主要来自布鲁塞尔自由大学的一群研究者，创造了一套20世纪的热力学。由此我们可以借助“自组织”这个新的科学法则，来理解秩序为什么可以在混乱中出现。他们的论点和通常对第二定律的理解不同，说第二定律并不等于千篇一律地朝着混乱一直消沉。混乱固然可能是物质的最后状态，在时间终点的宇宙固然会是一片倾圮，但是第二定律绝不是说这个过程均匀地发生在空间的每一点、时间的每一点。

首先我们必须区别变化潜能已耗尽的“平衡热力学”和“非平衡热力

学”。一杯咖啡已经冷到室温，不能再冷的时候，应该用平衡热力学；当牛奶才加进去，还没有搅匀的时候，就应该用非平衡热力学。如果我们把第二定律应用在事物不断在变的实际世界，而不应用在僵化的热力学平衡状态，那么自组织就会很自然地从第二定律中产生。平衡与非平衡热力学之间的对比，犹如实存和将然之间的区别，犹如这句话里面的字和句尾的句号之间的不同。在所谓宇宙向热寂退化的过程中，我们可以找到许多自动产生秩序的出色例子。咖啡里面加牛奶，最后状态固然是那个常见的灰色浑汤，但是在达到那个状态以前，白牛奶在黑咖啡里排演了多少瞬息万变的旋涡花样和结构！

“化学钟”是自组织的一个实验例子。这是一种特别的化学反应，里面的颜色很有规则地变来变去，也会显出美丽的旋涡结构。为了保持花纹，化学反应必须不断地得到补充。它的组织也很特别：是连锁在一起的一串化学反应，牵涉到反馈，其中一个反应的产品又参加同一个反应，甚至做它自己的催化剂。令人惊讶的是，化学钟里成千上万不计其数的分子，好像都精确地知道彼此在做什么，它们好像能彼此“交换信息”。

这些概念对生物界的含义非常重要。对生物来说，变化完结的平衡态就是死亡。热力学提供了一套自然语言，用它可以描述生物学的过程：只有远离平衡的过程，变化才能发生。我们能活着，完全是因为存在着一个由许多精密和谐的节奏组成的复杂网络。以这些节奏进行的生物化学反应，跟化学钟属于同一类。参与这些反应的就是生命的精髓——链状遗传分子DNA和RNA，它们能间接地催化本身的自我产生。这样，在经典热力学和达尔文进化论之间的鸿沟上，非平衡热力学搭了一座桥。它可以大致说明，为什么在一个熵递增的宇宙里，像人这样具有极其微妙结构的生物，仍然可以出现。

更引人入胜的是关于时间的含义。像化学钟这样的自组织告诉我们，第二定律不仅提供了一个时间箭头，并且它里面也有我们到处可以看到的各种循环、各种花样的种子。时间的这两个方面都很重要。

时间的箭头代表时间的前进，每个时刻都带有自己的烙印。可是另一方面，就如噪声和音乐的区别在于拍子和节奏，因此要在服从同一规律的现象里找出花样款式，时间的循环模式是至关重要的。第二定律为这两个最重要的时间形象打下了基础。

然而，即使我们承认第二定律的时间之箭意义深奥、根基牢固，不可能是幻觉，我们的主要问题仍旧没有得到解答。我们如何能使这个不可逆时间和用“无时间性力学”描述的微观世界和谐一致？这个谜，玻尔兹曼只解答了一部分。本书第八章将讲述，答案可能从正在萌芽的一门新学科里得到。这门新学科就是自组织的姊妹科目：动力学混沌。

## 混沌与箭头

在有关不可逆过程的讨论中，混沌并不等于全盘的天下大乱，而是指一种奇妙的秩序。描述化学钟行为的方程，解起来可以得到各式各样的解，不仅包括自组织，而且有“确定性混沌”。这是一种似非而是、可预言的偶发性。在化学钟的情况下，混沌表现为一连串颜色的随机变化。它之所以叫做“确定性的”，是因为混沌学专家们从这种随机行为里面，整理出来了一个微妙的基层组织。一般相信天气就是由混沌主宰，所以天气预报短期有效，时间过长就不灵了。目前科学界掀起一股混沌热，到处找混沌，吉卜赛蛾数目的起伏、羊痫疯的发作以及许多其他的现象，从政治到经济，全是混沌研究的对象。

时间对称的牛顿运动方程里，也存在混沌。这项颇为惊人的发现，含义极为深刻。据说，自命为“混沌福音传道士”的物理学家福特（Joseph Ford）曾经说过：“一个大革命正在开始。我们对宇宙的整个看法，都会改变。”研究结果表明，在最简单如仅仅3个粒子相互作用的情况之下，都会出现混沌。可预言性、确定性，这些几百年传下来的神话，这一来再也站不住脚，一个钟表式的宇宙就更不用谈了。

在这个世界里，动力学混沌是主导，不是例外。过去是固定的，而未来是开放的；这样，我们重新发现了时间之箭。

我们终于开始理解，不但是复杂的体系，就是物理学中最简单的情况，未来都是开放式的。牛顿力学和量子力学跟时间之箭配合，从而使创造性的演化成为可能，这看上去离不开混沌。自从玻尔兹曼以来困扰科学界的一个难题，现在答案总算有了点眉目。

# 第二章　牛顿物理学的兴起：时间失去了方向

科学！时间老人的真正女儿！

——坡（Edgar Allan Poe）

《致科学》

伊萨克·牛顿于1642年圣诞日出生在林肯郡乌斯索普的一个小庄园里。因为过于早产，他成活的可能性极小；他的妈妈汉娜说，他简直可以放进一个1夸脱的罐子里。可是在他以84岁之高龄去世时，大街上举行了游行盛典，举办了露天表演，诗人为他写诗，雕刻家为他雕像，赞颂他的生平。伏尔泰（Voltaire）曾写道："他像一个国王那样被安葬了，一个为其臣民做过大好事的国王。"

牛顿了解运动的意义，从行星绕日的轨道到箭矢射中目标的径迹，他都搞通了。在大瘟疫盛行的时日，他发展了包括时间在内的，对宇宙的第一个主要的数学描述，其中融合了哥白尼（Copernicus）、开普勒（Kepler）的天文学思想和伽利略（Galileo）的新运动理论。许多人把1666年称为他的奇迹年，正是因为他在这一年中在数学、光学和天体力学各方面迈出了巨大步伐。其实1665年也同样是奇迹年。在他50年后的一篇记述中，牛顿描述他自己的成就时写道："所有这些都是在1665～1666年大瘟疫的两年间取得的。因为那些日子我正年轻力壮，富于创造，以后就再也没有像那时那样专心致志于数学和哲学了。"

毫无疑问，牛顿对于数学和物理学的贡献是无与伦比的，它们开辟了一条分析物理世界的全新途径。他通过主宰天空和地球的定律而揭示了自然界的统一性。他一生的成就，被古体诗大诗人坡普（Alexander Pope）在西敏斯特大教堂著名的牛顿墓志铭上，简洁地概括为：

大自然

和它的规律深藏在黑夜里。

上帝说，

让牛顿出世吧！

于是一切就都在光明之中。

1686年4月28日牛顿将《原理》的第一册提交给皇家学会，这是物理科学的转折点。有些人认为《原理》一书是有史以来最伟大的科学著作，是科学文献皇冠上的一颗宝石。另一些人把它比作高耸于周围一群摇摇欲坠的临时建筑之中的一座大厦。牛顿也认为这书是自己所有著作中最高的成就。在其书第三册的开始他自豪地写道：“我现在就来说明世界体系的框架。”

时间在这个首次对运动的科学描述中肯定地出现，就像最早的钟表利用运动把时间的流逝转换为容易测量的量，例如钟摆在空间的摆动一样。在他的《原理》中，牛顿提出了运动的三个定律，这三个定律改写了关于运动的科学。他证明了地球上和天空中的物体都是被同一种力——引力——所支配，即使行星保持在一个轨道上，而像苹果这样的物体却落向地面。这样，他也同时解决了人类自太初以来一直大惑不解的一个问题——行星在太空的运动——这也正是守时和航海的关键问题。

为了把天空作为一个精确的时钟或历表，不仅需要有关太阳和恒星运动的信息，而且也需要有一套把这些资料编制成理论的方法。牛顿推导出天体运动的数学表达式，其精度是前所未有的。说起来也很妙，虽然牛顿对于天空的迷恋根源于神学，他的思想却成为人类斩断这种联系的铁砧。恰恰正是天体的运动、它们的轨道和重复的循环，最终导致了人们把理解的重心从魔法、巫术和诸神转移到科学和数学原理上去。正如罗素（Bertrand Russell）所写的："几乎所有现代世界与古代世界之间的区别，都得归功于在17世纪取得最辉煌成就的科学。"

时间的概念由人类对天文学的追求而出现的这一过程，鲜明地表现了科学思想的进化。古代的民族为了预期何时洪水发生，为了知道冬季的开始和第一个春日的到来，认识到了标志时间流逝的重要性。我们今天用时钟的指针在24小时的周期中扫过钟面来量度时间，而他们靠的是太阳的升起下落、月亮的周月运动或者载运众星的夜空从地平线的一方到另一方的转动。因此几乎所有的文化之中都有天文学的脉络，也就不足为奇了。

生命是围绕着维持生命所需要的能量而构成的，这种能量从太阳源源不断地流向地球。这座天钟的节奏对地球上的生命是如此之基本和重要，以至于几乎每一个活着的生物身上都反映着这种节奏（我们将在第七章和附录中再回到这个话题）。能量倾注到地球上的每一个角落，从太阳自东方升起之时一直到夕阳西下。这是一个由于地球绕着自己的轴转动而引起的视运动，它可以用来守时——牛顿据说就能根据影子的位置来指出一天中的时间。

太阳的第二个视运动提供给我们一个季节变化的时间单位——年——它大约等于365.25天。因为地球绕日的轨道周期并不正好等于整的天数，所以每四年要多加一天，也就是闰年，来防止误差的积累。太阳在天空中位置的季节性变化以及季节本身的出现，是由于地球在空间中的取向：它的自转轴相对于它绕日的椭圆轨道平面是倾斜的。

婆罗洲的部落利用观测太阳高度的变化来观察季节的更迭。在英国索尔兹伯里平原上矗立着一个巨石群，这是一座巨石堆砌而成的庙宇，古人用其石柱相对于太阳的列向而测定季节。另外还有一些天体时钟也曾被使用过。其中在古代计时中最重要的要算是月亮的周期了，每相隔大约29.5天，新月就在西方傍晚天空出现一次——这一运动周期近似等于我们的一个月。天上的星座也同样关联着季节的循环，有些星宿在夏季主宰着夜空，而另一些却在冬季出现。

天文学在人类历史上的起源要早于其他自然科学，它起源于史前时代，但最初记录没有被保留下来。在最古的文明中，占卜或占星术被用来解释恒星和太阳的运动。例如中美洲的阿芝特克族就相信太阳必须由血和一颗跳动着的人心来滋养，否则就会消失。在那些黑暗的日子里，科学与宗教之间没有冲突：教士、术士或沙曼（shaman）总是多疑善防地守护着关于季节和历法的知识。这些知识被当做是人世间神绩的标记，教士们因此在社会中拥有崇高的地位，因为他们能够预言未来，并取得了某些成功。天文学意味着凌驾于他人之上的权力。天文学通过季节指示人们，何时应当耕种，何时应当收获或者迁移牧群。宗教和祭祀活动同样也必须在特定的时节举行，例如与月相相合的日子或冬夏至日。天文学也帮助指引行路人。难怪《圣经》中跟着一颗星而走到伯利恒的三个人被认为是智者。

历法的兴起使人类的活动能更精确地与季节配合，因而更加协调。历法中所使用的三种周期——日，月，年——都是基于对人类生活具有最大的影响的天文周期。最早的历法依赖于月亮，因为它不仅有升起下落，还在一个月的周期中改变着位相，从而便于描述季节。后来的历法发展是根据太阳的周年循环。古代埃及人所使用的历法，被认为是古代最先进的太阳历之一。尼罗河的泛滥是古埃及人一年中最大的事件，而预言这一事件的关键是天文学，因为这一事件恰好与天空中最亮的星——天狼星（Sirius）——黎明前在东方地平线上出现相偶合。这一事件对古埃及人是如此重要，以至于他们把天狼星的

升起称为“一年的开启”，他们的历法就是围绕这一事件而编制的。

许多古代历法中使用基于新月之间大约30天平均周期的12个月，这样的一年太短，需要延长。最初的埃及历法就是用12个月，每月30天，这样一年就只有360天。后来在每年的末尾加上了5天来使得“太阴月”与基于太阳的季节保持协调，从而和“一年的开启”步调一致。

把从日出到下一次日出的一天划分为24小时的办法出自埃及人。他们用一个小时的间隔来标记从东方地平线升起的恒星或者星群，这样恒星横越天空的运动也就是12个小时，因而就有了12个小时的夜晚。后来很可能是为了对称性的缘故，也就出来个12个小时的白天。他们用水钟来测量白昼的时间，水从一个石制容器的孔中不断滴出而记录时间的流逝；他们还用日晷和影钟，利用阴影扫过钟面来显示时刻。但这样显示的是“不均匀时刻”，它们并不均等而且随着季节变化。在日本，这种“不均匀时刻”直到19世纪还在用，而且机械钟还要调得与之相应。在欧洲，从14世纪城市采用了机械钟后，一天就被划分为均等的24小时了。

我们现用的历法是由古罗马人使用过的历法而衍生来的。古罗马人用的是太阴月，为了补足太阳年，他们不时地插入一个闰月。到了恺撒（Julius　Caesar）时代，这种处理办法已混乱到冬天的月份落到了秋季的地步。这个置闰法被某些教皇和某些有权决定闰月的官员滥用过，这一班人为了政治上的原因用此延长公职任期或提前进行选举。到了公元前47年，这个历法和太阳年之间已脱节达三个月之多。次年，在希腊天文学家索西杰尼斯（Sosigenes）的指导下，恺撒不仅加了一个惯常的23天，而且还插入了两个追加月，使得那年的天数总共有445天。这一年后来被称为“混乱年”。从那以后，12个月的每一个月便具有现在的天数。

可是，一年365.25天的儒略历，每年还是要多出11分14秒。随着世纪的推移，历书上季节日期不断前移，恺撒时代发生在3月25日的

春分到了1582年已移到3月11日。那一年，教皇格里高利十三世（GregoryVIII）颁布了一种新的更精密的历法，同时把10月4日后面的一天指定为10月15日。然而新教徒不情愿接受天主教的这项革新。在英国直到1752年才用格里历来取代儒略历。随后一个姗姗来迟的11天的改正还引发了伦敦和布里斯托街头的骚乱，而使一些人丧生。工人们要求那几天的工资，很多人认为他们失去了自身生命的一部分。这一件事也影响到牛顿的出生日，这一日子按现代的格里历应是1643年1月4日。希腊正教直到1924年才采用格里历，但还有一些定期集市和地区性节日仍然沿用儒略历。穆斯林们用一种月亮历，因此他们神圣的斋月每格里年都得提前。

星期并不是根据天空的运动。正如伦敦社会研究所所长杨（Michael Young）所指出："太阳并没有成为唯一的主人。人类可以创造他们自己的周期，并不是必须依赖现成的东西。没有其他任何生物，能表现出对于天文学如此强的独立性。没有其他任何生物有星期。"星期的出现，很可能是因为社会对一个小于月而大于天的时间单位感到需要。如果人们洗衣、做礼拜和度假能有规律，社会活动就会进行得更加平稳。古代哥伦比亚人常用为期三天的星期。古希腊人喜欢十天一周而某些原始部落却偏爱一星期只有四天。七天一周源出于巴比伦人，后来影响到犹太人（虽然前者是以"恶日"而不是以"安息日"来结束一周，当天为了讨好诸神而施行各种禁忌——这也许是对星期天活动诸多限制的起源）。七天一星期广得人心，许多想改变它的企图都没有成功过。法国人在大革命后曾试图把它变为十进制，然而他们的十天一周终被拿破仑废弃了。在1929年（前）苏联曾尝试引用五天一星期，并在1932年把它延长到六天，可是到了1940年还是回到了七天一星期。

正如星期不理会天文学一样，现代守时技术也不理会天文学。现代科学的发展，使时间量到越来越小的间隔。像星期一样，小时被分成分和秒也起源于最早的科学天文学家巴比伦人，当他们在大约公元

前1800年完成他们的星表时，所有的计算都是以60进制进行的。但是只是到了工业革命以后，由于火车时刻表和其他一些详尽的工作程序表的需要，“分”才日益变得像今天这样重要。这种把时间不断越分越小的趋向，是和科学的发展需要处理极端快速的过程联系着的。例如，一个激光脉冲可以被用来捕捉类似原子在一个衰变中的分子里运动这类事件，这种事件持续的时间只有几十个亿亿分之一秒。

## 希腊人，古代天文学和科学

就牛顿的工作而言，最重要的天文学遗产来自古希腊人。古希腊人不仅像其他古代文明一样，收集了他们周围世界的信息，而且他们也试图用理性来了解宇宙的运作，而不求助于神、巫术和迷信。

为了解释天空的奥秘，牛顿不仅需要他的力学，而且需要一个天空的实际模型。生活在埃菲索斯附近迈勒图斯的塔里斯（Thales，大约公元前625～前547年）常被称为是最早的哲学家。他相信世间只有一种基本物质——水，而且认为地球是在球形的宇宙中的水上漂浮着。这种对称的宇宙模型以后还要不断出现。下一个有影响的见解是公元前6世纪毕达哥拉斯（Pythagoras，大约生于公元前560年）提出的，他是一个饶有情趣的古典嬉皮士组织——即毕达哥拉斯学派——的领导人。他们不喝酒，不穿毛皮制品并且食素。他们相信灵魂可以离开躯体。同时他们对于数学持有强烈的信念，他们挚爱对称，并认为“数字是实在的精髓”。根据他们对声学和行星轨道运行时间的数学知识，他们认为天空具有音乐的韵律，即所谓的天球音乐。这种观察事物的方法给予了天文学的发展深远的影响。甚至于牛顿也被说成是17世纪的毕达哥拉斯，“他把一生贡献给对宇宙和谐的研究”。

这种圆周运动的对称性，例如一个转动的轮子所表现的，是古代宇宙模型的中枢。这种圆周模式影响了人类工具的发展50万年，现在它开始影响作为思维工具的理论模型。圆是最完美的曲线，因此它也

成为描述行星如何绕着地球运转的、美学上最具吸引力的模型。也由于这种对于圆周对称性的喜好，古代人认为天空和地球必须是圆的。

公元前400年左右是古希腊思想史的黄金时代，那时柏拉图（Piato，公元前428～前347年）的思想地位最高。像毕达哥拉斯一样，柏拉图也赋予数学极重大的意义。遗憾的是，他的知识论的基础是，我们观察到的世界并不像实际的世界。他认为，一个完美的宇宙模型，主要是要显示神的尽善尽美，而不是要描写我们的观察。他的模型很多与毕达哥拉斯有关（可是与菲洛劳斯（Philolaos）无关——他是毕达哥拉斯的门徒之一，认为地球像其他行星一样也在作轨道运动）。对柏拉图而言，实际世界的观测和实验与知识的探讨无关：真实的实在只能用头脑去深思而得。在他的宇宙学大作《狄玛尤斯》中，宇宙是一个井然有序的世界，地球位于中心，其他的天体在不同半径的球面上运动。无疑地，尽管柏拉图有对数学的癖好，或者正是因为如此，他对实验方法的不喜欢严重地阻碍了科学的发展。

公元前384年亚里士多德（Aristotle）的诞生带来了希腊科学的新纪元。作为柏拉图的学生，他也认为地球处于宇宙的中心，其他的行星在不同的球面上运动，而恒星镶嵌在最外面的球面上。与前人相比，亚里士多德赋予观测的地位要高得多，这为现代的科学工作奠定了基础。他提倡观测与理论之间的相互影响，观测显示世界的运行方式，而理论解释其原因。但是从我们后人的观点来看，亚里士多德使科学的发展倒退得比柏拉图还远。这位西方哲学之父宁可用目的论的链条来解释世界，而不用因果关系。宇宙目的论（Teleology）探求的是宇宙现象的目的。如果要我们解释座头鲸的存在，我们会援引达尔文的进化论，一种因果论证，而宇宙目的论却把这归因为仁慈的造物主（上帝）对人类的恩赐。也许宇宙目的论的最著名的例子就是“设计论”了，它被许多宇宙目的论者作为上帝存在的证据——特别著名的是帕雷（William　Paley）1802年写的《自然界的宇宙目的论》一文。在这篇论文里，帕雷满怀激情地争辩道，生命组织是如此之复

杂，因而它必须要有一个设计者——上帝。

今天这种论证方式已被看做是本末倒置的科学的神圣化。达尔文进化论的现代鼓吹者道金斯（Richard Dawkins），在他的《盲人钟表匠》一书中，先把帕雷的方法论优美地表述出来，然后说它是一文不值。然而令人玩味的是，这一设计论在今天，又被一些天文学家和宇宙学家以一种所谓人择原理（The Anthropic Principle）的极端形式重新提了出来，关于这个原理我们将在第三章中再遇到。简言之，人择原理认为宇宙之所以是我们所看到的这样，是因为如果宇宙不是这样的话，我们也就不会存在，也没有人会来观测宇宙了。

萨摩斯地方的阿里斯塔克斯（Aristarchos of Samos，一般认为生活于公元前310～前230年）第一个提出太阳中心说，他的天体模式与现代的观念是相一致的。他坚持地球是围绕一个固定的太阳做圆周运动，而不是位于宇宙的中心。但是他的思想由于缺乏经验根据，在亚里士多德派的影响之下，被打入冷宫达将近2000年之久。此后，阿奎那斯（Thomas Aquinas）研究了阿拉伯人保存的亚里士多德派手稿，由于他的支持，亚里士多德的模式得到了天主教会的珍爱，因为这一模式把人类放在宇宙的中心地位。15世纪中叶科学大革命开始，并一直进行到16世纪末。在这个时期，神的干预这一观念逐渐衰落，科学的宇宙思想代之而起。正如在牛顿出生那年去世的伽利略所说："圣经所指出的是通往天堂之路，而不是天空自己走的路。"

## 科学大革命

科学大革命代表了几方面思潮的同时繁荣，这些思潮可以上溯到古希腊时数学作为一门独立学科的年代。一些早期的天体模式，例如阿里斯塔克斯的模式，包含有地球绕日运行这种"现代"观念的胚芽。然而由于亚里士多德观念的影响，直到波兰教士哥白尼（Copernicus，1473～1543）的时代，没有重要的知识界人物认真

看待这个观念。哥白尼生活于文艺复兴时期，那时候出现了透视法的数学概念，给予艺术一个新的维度。同样地，哥白尼也改变了透视宇宙的方法，从太阳的角度而不是从地球的角度去想象它，从而解释了行星复杂的运行轨迹。

按照哥白尼的观点，太阳处于静止，而地球却被抛到空中。如他的1543年出版的《天体运行论》一书中所述，他把太阳置于行星体系的中心。一个世纪以后，所有能被接受的天体模式，都以太阳替代了地球，作为行星运动的中心。1729年去世的布莱克默爵士（Sir Richard Blackmore）曾写道：

哥白尼正确地判决了古老的体系，

创造了更完美的世界模式；

他叫太阳静止，

而令地球绕着它的真极运转。

然而，在一段很长的时间里，哥白尼的宇宙模型并没有被普遍接受。当时流行的观点是根据对圣经的字面解释，说人类占有中心地位。甚至在讲述新世界体系的《天体运行论》发表以前，路德（Luther）在1539年就曾抱怨过："这个傻瓜想把整个天文科学颠倒过来；但是神圣的圣经告诉我们说（约书亚记第十章第十三节），约书亚是命令太阳而不是命令地球停下来。"奇怪的是，直到布鲁诺（Giordano　Bruno，1548～1600）作为一个新学说的热烈鼓吹者、积极地传播它以前，罗马天主教廷并没有对哥白尼的新学说横加摈弃。其后哥白尼的思想被搞得与布鲁诺所倡导的与其说是科学不如说是巫术的那一套，混淆在一起。布鲁诺被天主教宗教法庭监禁，最后被烧死在火刑柱上，随之哥白尼的理论也就成了罗马天主教廷的攻击目标。这之后主要的贡献出自开普勒（Kepler，1571～1630）和伽利略（1564～1642），这些贡献的结果是：出现了一个结合时间的

数学框架。伏尔泰写道：“开普勒之前，所有的人都是瞎子。开普勒有一只眼睛，牛顿有两只。”开普勒在他的导师，最伟大的肉眼观测天文学家——第谷（Tycho Brahe，1546～1601）——去世前一年移居到布拉格。第谷临死前在床上嘱咐开普勒去编制一个行星运动表，以支持第谷本人的理论（与哥白尼的相反），即地球仍旧牢牢地位于所有天体的中心，月亮和太阳在环绕它的轨道上运行，尽管他也承认行星可能是绕着太阳转动。

开普勒的工作表明第谷和哥白尼都不对。在他的《新天文学》一书中，他摈弃了源于古希腊人的传统观念，明确地显示了火星不仅绕着太阳运转，而且它的运动速度老在变，轨道还是一个椭圆。这一来，行星运动再没有神一般的完美性——亚里士多德伪科学的棺木上又钉上了一根钉子。开普勒用三个定律阐明了行星的运动：在第一个定律里他描述了轨道的椭圆形状；第二个定律说明了行星的轨道速度如何变化；第三个定律给出了轨道大小和轨道周期之间的关系。但是在他关于世界运行的观念中仍然有着神学的烙印和神秘的色彩。例如在他的《宇宙的和谐》一书中，他把行星的最大和最小的速度联系于音乐的和声——令人重新想起毕达哥拉斯的“天体音乐”。

现在我们来谈现代科学研究方法的鼻祖——伽利略。伽利略是一个音乐家兼音乐史家的儿子，与莎士比亚同年，即1564年出生于意大利。伽利略对现代动力学科学的贡献，我们不可能低估。他系统地阐述了“加速度”的概念，而牛顿的第二定律就是建立在这一概念上的。“速度”的概念，作为量度一个物体的位置随着时间变化快慢的量，当时已经是家喻户晓。可是加速度内含更多的时间成分：一个运动物体的加速度需要我们找到在固定的时间间隔内，速度在方向或大小上的变化。伽利略强调，一个运动物体只有在受到某种力的作用时才会发生速度的改变，才会有加速度。任何一个曾试图端着一满杯咖啡在一列行进的火车的通道中行走的人，都会亲身体验到力和速度改变之间的关系。

伽利略第一个把望远镜应用于科学研究，从而把哥白尼的革命置于坚实的实验基础之上。望远镜的发明者究竟是谁，这一问题今天仍然是学者们争论的一个问题。但伽利略已能用之来注视现实宇宙的缺陷，而不再管古希腊的伦理偏见。伽利略1609年一听到荷兰的利普谢（Hans Lippershey）有关望远镜的工作之后，在几天之内就造成了他自己的仪器。第二年，在《星辰信使》上，他成为第一个发表望远镜观测结果的人：“啊，有了这样绝妙的仪器，新的观测和发现会有止境吗？”从木星诸卫星的运动变化，金星的位相以及太阳本身的转动，他亲眼看到了明显的证据：哥白尼的的确确是对的。他本来早已是一个哥白尼派；这时候他把每一个发现都用作捍卫日心说的武器，例如，他用木星被其卫星环绕作为太阳系的一个比喻。

但是他的工作使他与教会直接发生冲突。1952年爱因斯坦在为伽利略的《两个主要世界体系的对话》一书所写的序言中写道，伽利略“以其激昂的热情、智慧和勇气，作为理性思考的代表，挺立在教士们的迷信面前”。他还写道，伽利略工作的主题是，“反对任何基于权威的教条，坚决全力奋战”。

因为传播与天主教会教义相抵触的思想，伽利略被抓进宗教法庭。1633年他两次受到严刑拷问的恐吓，尽管他已宣布放弃了他的观念。教廷对他说，教廷“强烈地怀疑他持有异端邪说”，但只要他以“一颗真诚的心”“恳求并诅咒和痛恨上述的错误和异端邪说”，他就可以得到宽恕，而只被监禁。当他在佛罗伦萨附近的阿西特里的家被软禁时，他已69岁高龄，但仍继续在做力学方面的研究。同时他在对时间的科学阐述方面，也扮演了另一个重要角色：在他生命的最后日子，他投身于守时研究和利用钟摆来控制时钟机构。他的实验结果，荷兰科学家惠更斯（Christiaan Huygens）在1656年开始实用化，这样一个精确守时的新时代终于来临了。

当伽利略的《世界体系》一书流行于整个欧洲的时候，新教已经在几个国家确立，并与天主教会分庭抗礼。在这些国家，政府基本上

和教会是分开的：不管新教教士多么反对哥白尼，他们没有权利去压制哥白尼的学说。

## 牛顿和力学

古希腊人懂得静态的实体，例如各种几何体，但他们对于运动缺乏清楚的了解，例如，一支箭在飞行的过程中，是如何从一个时刻到下一个时刻改变它的位置的。这并不奇怪，因为要研究物体在自然界的运动就需要一个好的钟表，而他们当时缺乏这一基本技术。根据伽利略建立的基本原理——钟摆可以作为可靠的守时工具，彭罗斯（Roger Penrose）写道："牛顿得以建立起一座宏伟壮丽的大厦。"

牛顿的《原理》一书写于1684～1687年。有些人把牛顿看成是一个集大成者，把哥白尼、开普勒的天文革命观和伽利略、笛卡儿的运动理论结合在一起。另有一些人则认为牛顿学说中，有很多的新见解，所以他是个道地的创新者。尽管如此，牛顿自称他的思想源于古人，"主要源于塔里斯"。他认为毕达哥拉斯的天球音乐里面已经蕴藏着他本人的引力平方反比定律；如上所说，这个天球音乐思想也曾启发过开普勒。有人认为，牛顿这样引用古人，是为了提高自己思想的地位。

牛顿的《原理》包括一篇引言和三册书（或三个部分）。引言中写有如下的运动定律：

1.一切物体保持它的静止或匀速直线运动状态不变，除非有力加于其上迫使它改变这一状态。

2.运动的改变正比于所受到的动力，并且发生在该力所施的方向上。

3.对每一个作用都总有一个等量的反作用：或者说，两个物体的

相互作用彼此依赖，并总是大小相等，方向相反。

从他在书的前两册中所建立的数学框架出发，在第三册中他推导了行星、彗星、月亮和海洋的运动。在牛顿的框架之下，只要假设行星的加速度反比于行星到太阳的距离的平方，开普勒的行星轨道定律就可以很自然地推导出来。

一个行星能绕日运行，一定是受到某一力的支配，才把自己保持在绕日的轨道上，因为尽管它以不变的速率运动，运动的方向却是时时刻刻在变的。牛顿认为行星和太阳之间的“力”是速度改变的原因，具体地说，他认为加速度是正比于这作用力的。他著名的引力定律公式说，两个物体，例如太阳和月亮，它们之间的引力正比于它们各自的质量的乘积，而反比于它们之间距离的平方。换句话说，如果其中一个的质量加倍，则引力也加倍；但是如果距离加倍，引力则变为原先的四分之一。

伏尔泰说过，“伊萨克·牛顿爵士在他的花园里散步，看到一个苹果从树上掉了下来，从而首次想到他的引力理论”。按照《牛顿手册》所载，看来牛顿至少也用过这个故事来形容引力，但是说他曾被这个苹果打中，并且就由此而产生了万有引力的想法，这些都是夸张。只有一件事可以说是毫无疑问，就是在沃尔斯索普的确是有这么一棵树，它结出一种名为“肯特之花”的烹调苹果，缺少味道，样子像梨，红色中带有黄、绿条纹。虽然这棵树1828年曾被吹倒，但它的后裔仍在繁衍，因为它已被接枝到贝尔顿地方的贵族——布朗罗（Lord Brownlow）的一些树上。

为了从他的定律中得出肯定的预言，说行星和苹果应该怎样运动，乃至老鼠如何从柱子上滑下以及其他一些动力学现象，牛顿发明了微积分，这是数学的一门分支，至今仍是高等数学和理论物理的基石之一。这也引起过科学史上一次最著名、最激烈的争论，因为牛顿和德国哲学-数学家莱布尼兹（Gottfried Wilhelm Leibniz，1646～

1716）都声称自己发明了微积分。莱布尼兹很多年来都直截了当地在说，牛顿的学术研究大有问题，这无疑使争论更加白热化。1711年，莱布尼兹曾向伦敦的皇家学会申诉；两年以后皇家学会做出了一个有利于牛顿的“公平”裁决。很多年后真相暴露了，原来作为皇家学会会长的牛顿本人就是这仲裁书（Commercium epistlicum）的作者之一。威斯特福（Richard Westfall）写的牛顿传记《从不休止》中，认为这是牛顿的典型的性格：冷酷、自私、骄傲、惯于欺骗。“伊萨克·牛顿为人令人讨厌”，宇宙学家霍金（Stephen Hawking）也这样承认道。霍金是剑桥大学著名的卢卡逊教席的现任教授，当时牛顿就是该教席教授。

微积分是一种优美简洁、描写宇宙中许多事物的方法，但这里我们不想讲它所牵涉到的数学，我们只想用文字来描述一下，时间是怎么样在牛顿的方程出现的。如上所述，大到行星、小到甲虫，任何物体的运动都涉及物体的位置随着时间的变化。牛顿认为时间是一个绝对量，但他对时间的态度是要应用它，而不是要描述它。牛顿首先得解决一个问题，即如何精确地描述一个物体的运动。如果一辆马车在100秒内跑了200英尺，那么它的平均速度是每秒2英尺。但是假设有人要问50秒后跑得是多快。在那一瞬闻——其持续长度等于零——马车明显的是一点都没有动。问马车在一刹那间的速度，好像在问在一个短到运动都停止了的时间间隔内跑了多远一样。

牛顿解决这个问题的办法是，考虑一个物体在一系列很短的时间间隔内的运动。一个数学家也好，一个拦路强盗也好，如果给他提供马车运动的所有细节，包括转弯时走多快，什么地方慢下来等，他就可以估计出什么时候马车会通过一个预定的地点。只要用一个钟表，一个人就可以测出马车每几秒内走的速度。他的测量原则上可以推到一个极限，即马车在一系列越来越短的时间间隔内的速度。当时间间隔趋向于零的时候，这个速度极限就是瞬时速度。在这样无限小的时间间隔内，马车位置的变化也将是无限小的。然而马车的这一瞬时速

度却是一个有限的量：它等于微小的位置变化除以同样微小的时间间隔。在微积分学中，这个量被称为一阶导数。用这个极限过程，得出一个瞬时速度的精确描述，这是微积分的基本法则。同样地，用一个类似的极限，我们可以得到马车的瞬时加速度。

牛顿把这一新的数学方法运用于天文学。他看到行星围绕太阳的轨道被引力弯曲成椭圆形。这样，一个围绕太阳运动的行星其实是在不断地下落。它的轨道的形成，是由于它所受到的引力每一瞬时都在改变着它的速度。为了严格地用微积分来计算行星轨道，我们必须知道每一瞬时的准确加速度，而不仅是一个粗略的平均值。因此，瞬时加速度就在牛顿运动定律的数学公式中出现，这些运动定律其后被用于对太阳系运行日益详尽的描述。这样，用诺贝尔奖获得者温伯格（Stephen Weinberg）的话来说，“牛顿破除了两种形式的物理之间的壁垒。他不仅消除了太空的神秘性，而且开辟了把天空和地球放在一起研究的可能性”。

牛顿对于运动的描述，把人类对于宇宙结构以及时间的看法全部改头换面。罗素甚至认为，由于牛顿科学的兴起，“在1700年，受过教育的人的思想境界就已经完全现代化了；而在1600年，除去极少数以外，绝大多数人仍处于中世纪。”神在世界秩序中的作用被大大降低：上帝，即使他存在的话——而拉普拉斯（Pierre Simon de Laplace，1749～1827）和其他一些人认为连这假设都没有必要——也只是在时间的开端给所有的物体一个推动，其后就再也不需要他管了。实际上当时根本不知道时间是否会有一个开端。总之，天体的运动成为可以预知的了。虽然人类在宇宙中的地位变得无关紧要，但是人们开始赞美人类聪明才智的力量。不可避免地，这导致了神学上对上帝和人类关系的重新评价。教会的教条几乎是黯然失色了。

然而牛顿意识到他自己工作的局限性。他写道：“我不过只是像一个在海边玩耍的孩子，以偶尔间发现了一个更光滑的卵石或者更漂亮的贝壳为乐，而我面前仍是一片未知的真理大海。”（这个独特的

比喻很可能是二手货；牛顿从来没有去过海上，甚至于连去海边散步都没有过）实际上，牛顿在其自然观上是与上帝合作的，因为他有着强烈的宗教信仰。他坚持认为太阳系需要上帝时时刻刻的照料，否则就会不稳定。对于牛顿来说，上帝是宇宙的造物主和维护者，但这却使莱布尼兹讥讽牛顿的上帝好像一个二等钟表匠，造出来的钟表每次停下来还非要他自己去维修不可。

为了检验他的新理论的预言，牛顿利用了弗雷姆斯蒂德（John Flamsteed，1646～1719）的月亮运动的观测结果，后者是格林尼治天文台的首任皇家天文学家，也是牛顿为了知识产权问题后来与之争吵的另一位科学家。格林尼治天文台当时除了做纯研究外，还在另一项工作中起着重要作用：为使船只在海上能够找到所在的经度而做天文观测（纬度可以通过测量恒星的地平高度而得到）。尽管弗雷姆斯蒂德和他的后继者作了大量观测，这一工作直到1767年才算完成。大约也是在这个时间前后，来自亨伯河边巴洛城的一个木匠哈里森（John Harrison），制成了第一台天文钟，它可以在海上显示“家乡时间”。但只是到了下一个世纪，在1884年，才把这个“家乡”确定在格林尼治；从此，为使全世界的守时标准化，格林尼治时间建立了起来。比较格林尼治时间和本地时间就可以得到经度。比如，如果海上的日出时刻比格林尼治时间晚四个小时，船长就知道，他已经绕地球航行了六分之一。

## 牛顿方程式的剖析

牛顿的运动方程式把一个物体的加速度直接与作用力联系在一起，这样就产生了一个有关时间的奇妙结果。在加速度的瞬时值中时间出现两次：加速度是速度随时间的变化率，而速度是位置随时间的变化率。在微积分中，这样一个量被称为位置对时间的二阶导数。这就在牛顿的运动方程中产生了一个重要的后果，即时间是作为两次幂，也就是平方出现的。如果我们把前进的时间（“正的时间”）替换

为倒退的时间（“负的时间”），这些方程不会改变，因为两个负数的乘积和两个正数的乘积一样总是正的：负时间的平方等于正时间的平方。这样，牛顿力学是不能够区分这两个不同的时间方向的。牛顿方程式本身不能告诉我们是在变老或是变年轻，但这一点正是人生至关重要的一个方面。

一个球从球拍上弹回也好，水星在绕日轨道运行也好，对牛顿方程的每一个解，只要把时间方向简单地颠倒一下，就可以得到另外一个同样可以被容许的解。这等于设想时间倒转。

我们马上会觉得，时间对称是一个非常特殊的性质，因为我们直接经验的现象中，这种情况极为罕见。考虑图1所示的两个行星围绕太阳运行的情况。（a）和（b）是不同的运动状态，但都是牛顿方程式同样容许的解。这样，记录（a）的电影胶片如果倒过来放映就会变成（b）；但不论这电影如何放映，都不能告诉我们哪一种情况代表时间真实的方向，哪一种情况代表时间的反向。我们可以说，牛顿运动方程式描述的是一个完全可以逆转的世界。

然而，这是我们所了解的那个世界吗？试想公牛闯进瓷器店的影片（图2）。滑稽的时间倒转——被打碎了的瓷器奇迹般地重新聚合在一起，公牛倒退跑出这家商店——在现实世界中从来不会发生。这两种过程中哪一种代表了时间流逝的正确方向，是不证自明的。

还有无数其他的例子：从来没有人见到过一杯茶会自发地变热——它总是自己凉下来。我们只见过季节以同样的春、夏、秋、冬的顺序出现，而从没有见过夏天紧接着秋天。我们与单向过程最直接的接触大概就是年龄的增长。从来没有一个记录在案的例子，显示一个死亡了的有机体会复活，越变越年轻并最后“倒生”回去。像生命一样的单向过程，叫做不可逆过程。

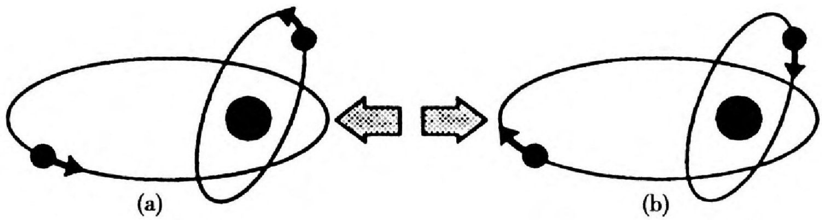


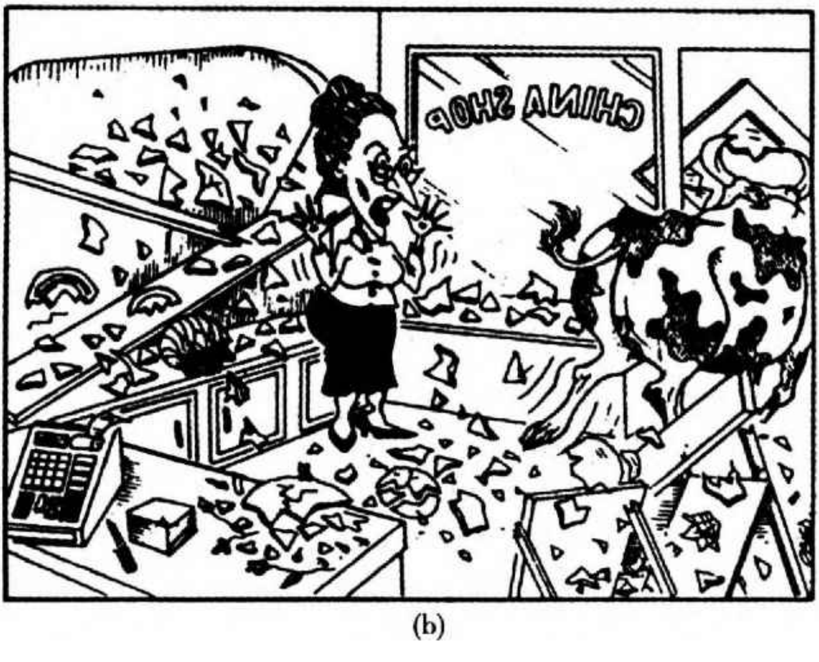

图1　牛顿力学的可逆性。在时间向前和向后之间没有区别。图（a）表示围绕太阳运行的两颗行星。如果把记录A的影片倒放，我们就会得到图（b）。但是我们如何才能独一无二地确定，哪一个真正是时间在向前走呢？[录自柯文尼，法文杂志《研究》，第20卷，190页（1989）。]

然而牛顿的定律显然预言，这样不可能的时间倒转过程是完全可以发生的。公牛可以溜达走进摧毁殆尽的瓷器店，把破碎的碗碟拼起来并重新堆放好。是不是我们因此就可以得出结论说，牛顿的运动理论是不正确的，因为我们已经看到这么多的实际过程，它们看来都与牛顿定律的预言相抵触？

为了避免这一结论，有人就说，应当考虑这个故事实现时间倒转所需要的初始条件。公牛把瓷器碎片拼合好原则上是可能的，但是可能性极小极小。把一个盘子打成碎片可以有许多方式，但是把这些碎片重新还原只有一种方式。用牛顿方程式去拼合瓷器碎片，初始条件实现的几率是如此之小，所以我们看不到改过自新的公牛，去重整被糟蹋了的瓷器店。这样，时间箭头便出现了。




(a)




(b)

图2　在真实的世界中我们决不会遇到完全可逆的系统。日常发生的一切都是不可逆的。图中所画是想象一头公牛传入一家瓷器店的情况[（a）是闯入之前，（b）是闯入之后]。从来没有

见过（b）发生在（a）之前的时间倒转过程。[录自柯文尼，法文杂志《研究》，第20卷，190页（1989）。]

但这个说法是有缺陷的。根据这个说法，时间的箭头不是一个内禀的性质，而是由于这家瓷器店起初特别整齐有序。这类似于说，一个放在斜坡上的球将总是向下滚动。对于这个说法，牛顿的时间对称的方程式仍然可以作如下反驳：一组初始条件，既可以作为时间向前的事件的出发点，也同样可以作为时间向后的事件的出发点。反过来说，如果起始条件能够任意选择，那么一个在山底的球也会滚上山顶的。

也许这时间箭头出自于耗散。耗散是区别可逆过程与不可逆过程的一个关键特征，它涉及所研究的系统中的能量再分布，或能量从该体系的散失。瓷器被冲撞时，“摩擦力”或者其他的力起作用，使能量耗散，能量在瓷器碎片和周围环境之间转移。我们知道，碎片之间的相对运动由于摩擦力而减慢而不是加快，因为部分动能被转换为热能。这种摩擦力的“衰减”作用可以被加进牛顿方程式，只要简单地承认它的存在，在方程中加进一个附加项来代表它就行。原方程的时间对称性因此被破坏，不可逆情况于是出现。

然而基本的困难并没有解决，因为根据现代原子论的观点，摩擦仍然需要用原子和分子的运动来解释，而这种运动本身是服从牛顿可逆的定律的（或者同样地，服从第四章要讨论的“无时间性”的量子力学定律）。因此，我们只是把问题暂时放在一边而已。牛顿的方程与时间方向无关，一只摔碎了的茶壶，它的几十亿分子原则上可以开倒车，从声波中吸收所需的能量，使散落的瓷器碎片自发地重新拼合起来。

也许牛顿力学并不是可以到处应用的，也许我们为解释时间之箭已把它用在不应该用的场合。但是，时间的箭头，确实是许多过程包括生命现象在内的一个内禀特征，这些过程并不特别依赖于“初始条

件”。我们将在第五章中讨论这些过程。

## 电磁时间

牛顿提出他的定律，是为了研究引力对大质量物体的作用。但是自然界中还有其他的力，例如静电力——我们梳头时使头发竖立起来的力。解释静电现象的定律最后演变为电磁理论，它是物理学的第二个主要的理论构成。在这里，时间同样是一个棘手的问题。

牛顿引力理论的一个饶有趣味的特点，是它描述了两个大质量的物体（例如太阳和月亮）之间的一种瞬时作用，尽管这两个物体并没有直接接触。这种现象被称为超距作用。它使当时的科学家和哲学家都感到头疼，因为找不到显而易见的机制去说明它。在《原理》一书中，牛顿叙述道：“我希望我们能用类似力学原理的推理，导出其他自然现象，因为有许多理由使我猜想，这些现象都取决于某些力，这些力使得物体中的粒子由于某些迄今未知的原因，或者相互靠近而连接成规则形状，或者相互排斥而分散。这些力既属未知，所以哲学家们迄今对大自然的探索仍是徒劳无功。”

对于像一记拳或是一记耳光那样的碰撞力，物理学家和哲学家们可以理解。可是对于吸引力或排斥力——像牛顿的引力——他们总认为是玄虚的。牛顿在科学上的主要敌手莱布尼兹，曾把牛顿的工作评论为“引力（不言而喻，任何牛顿其他的原动力），不是故弄玄虚就是某种奇迹的作用”。牛顿为了解决这一问题，想象了一个引力场，它从每一个引力质量中流出，瞬时弥漫到整个空间，并且随着到物体质量中心距离的增加，它的强度按平方反比而减少；这样当距离增加一倍时，引力场的强度就减少到四分之一。

静电力——例如，在带电的梳子和头发之间的静电力——以同样的方式作用于整个空间。为使这种作用在一段距离外发生，就要假定

有一个电场，就像牛顿的引力场那样。1785年法国人库仑（Charles Coulomb）获得了必要的实验精度，为静电力的理论提供了基础。根据他的实验，他得到了一个把荷电物体之间相互作用定量化的定律。库仑使用了一个扭矩天平，这是一个可以测量一对荷电球之间电力的装置。他发现同性电荷相互排斥而异性电荷相互吸引，在这两种情况下，相互作用力都准确地按照荷电球之间距离的平方反比而变化（并且正比于两个球电荷量的乘积）。

库仑定律与牛顿的引力定律具有惊人的相似性：两者都用了场的概念，都用了平方反比定律，来描述超距作用。诚然，也有一些重要的区别。电荷有两种类型，正电荷与负电荷。同性电荷相斥，异性电荷相吸。而引力只有一种类型的“荷”——质量——它总是相吸的：日月星辰之间全都是互相吸引。

与静电学有关的静磁场的研究，与静电场有非常相像的历史，两者之间有许多相似之处。当时担任伦敦皇家研究所所长的法拉第（Michael Faraday）1820年在电学和磁学方面进行了独创性的研究，发现运动的或动态的磁作用与静电作用紧密相关，而且反之亦然。运动的电荷产生磁场，而运动的磁场在导体中产生电流（第三章中我们将深入讨论这种对称性的原因）。法拉第的开创性工作，由苏格兰人麦克斯韦（James Clerk Maxwell，1831～1879）用有力的理论继续发展。麦克斯韦1864年当伦敦大学皇家学院的教授时，证明了电和磁的作用，是同一个电磁力不同的表现形式。他最后集其大成的数学方程是如此优美，使得玻耳兹曼（Boltzmann）不禁引用哥德（Goethe）的语句：“难道是上帝写的这些吗？……”麦克斯韦把法拉第的电磁定律数学化，其结果现在就叫做麦克斯韦方程。根据这些方程，麦克斯韦得到一项推论说，电磁信号在真空中应该以一个恒定的速度运动，而这个速度就是光的速度。

这样说来，我们就很难避免下结论，说光本身就是一种电磁作用。不久之后，另外一些形式的电磁辐射也被发现了，从此人们知道

可见光只是电磁波谱中的一部分，整个电磁波谱覆盖着从射电波直到X射线以及它以外的波段。我们熟悉的从红色到紫色的电磁辐射波谱，仅仅只是整个波谱中的、人的视网膜感觉得到的一个波段。

然而，就像牛顿方程一样，麦克斯韦方程也不区分过去和将来。时间不论是正值还是负值，方程都是不变的，方程里面不包含过去和将来的区别。按照麦克斯韦方程，一个像电子这样带电的、有质量的粒子，在电场和磁场并存的情况下，由于同时受到这两个场的作用，将受到一个以荷兰物理学家洛伦兹（Hendrik Lorentz）的名字命名的力。这个粒子的运动于是就可以用牛顿运动方程来描述，洛伦兹力和粒子质量决定粒子的加速度。

这样我们又一次失去了时间箭头。正如先前讲到的引力下的运动一样，现在我们在电动力学中又遇到了可逆的力学描述。有关带电粒子在电场、磁场或者两者并存情况下的实验，证实了这些时间对称的运动方程的解，的确给出了正确的动力学结果。可是许多电磁现象，很明显地是具有时间方向的。从没有人见过光波从照亮的房间里聚回到电灯灯丝，然后被灯丝吸收；也从没有人见过光线从我们的眼睛跑出来，再被太阳或是其他光源吸收回去。因此有些人说，存在一种电磁的时间箭头，它可以排除这些“倒转”过程，原因是这些过程的初始条件被实现的几率极小。这种说法和我们前面已经反驳过的，有关公牛和瓷器店的说法十分相似。

电和电磁辐射在守时技术方面起了很大的作用。依赖于个别地方准确守时的“地方时”制度由此结束，取而代之的是全国性的“国家时”。这给出了一个全国范围共同意识的“现在”。无线电波可以使遍布全国的钟表时间同步。当第一个电报系统1838年在英国被采用时，人们就已认识到，用同样的办法，可以传播来自同一个主钟的信号。电使得钟表的准确性比以往大大提高。在美国的贝尔实验室，借助于电路装置的晶体石英钟，早在20世纪20年代后期就已经问世。在这类钟里，石英晶体像音叉一样，以恒定而且非常准确的频率振荡。这一频

率是石英晶体的特性，与机械钟不同的是，它和钟表的设计基本无关。

1948年，设在华盛顿的美国国家标准局成功地把一种分子振动用于守时，为原子钟铺开了道路。原子钟的“滴答”频率是完全与工艺设计无关的。美国国家标准局当时用的是氨分子，它的形状像金字塔，由三个氢原子和一个氮原子组成。三个氢原子构成一个环，氮原子前后跳动穿过这个环，就形成了钟的“滴答”走时。最古老的守时钟就是我们所在的这颗行星了，它的缺点是它的转动速率不是完全稳定的，与此相比，原子钟要好得多了。由于地球极冠的冰雪冻结和融化，潮汐的摩擦以及其他产生于地球内部深处的作用，一天的长度在一年之中，有千分之一秒左右的涨落。这对于现代超精密的守时需要来说，是完全不够的。

## 关于场，以太，空间和时间

麦克斯韦所预言的电磁波，照他自己的描述，可以想象为“带电或带磁物体周围空间中，一种电磁场的扰动”。这种波的特征，可以用它们的波长来表示——即相邻两个波之间，振动相同点的距离。射电波的波长一般是几米或更长，而放射性原子射出的伽马射线的波长要小亿万倍——大约是氢原子直径的百分之一。我们眼睛的视网膜能感觉到的可见光，其波长是原子直径的几千倍，介于上述两种极端情况之间。

但是，载运电磁波的介质是什么呢？我们最熟悉的波，并不是从太阳晒到日光浴者头顶上的电磁波，而是从海里打在沙滩上的水波。难道麦克斯韦的电磁波，也一定要经过什么东西才能传播吗？麦克斯韦假定有一种叫做“以太”的介质，是它在载运电磁波。有趣的是，以太这个观念，起源于绝对空间的概念；直到爱因斯坦时，才把这个观念扫除掉。

绝对时间和绝对空间源于古希腊思想，特别是亚里士多德。牛顿和其他许多人构想过一种参考框架，一种绝对静止的状态，它类似于一个巨大的、延伸到整个宇宙的网格。相对于这个框架或状态，一切物体的运动就可以在实验上与理论上与它比较。例如，一个在火车上往餐车走的乘客，他的速度可以从每小时几英里变化到几千英里，这取决于他的速度是相对于火车、铁轨测量，还是相对于牛顿假设的绝对静止状态。正如我们已经提到过的，牛顿同样设想了“绝对时间”，它独立于空间，在任何地方都以同样的快慢流逝：“把时间联系于恒星运动（如柏拉图建议的那样），或者联系于‘运动数’（亚里士多德）、意识（奥古斯汀，Augustine）、世界和人类（亚维若艾斯，Averroes）或是生命和感觉，这些企图都已经一去不复返。时间成为一种普遍规律，不论发生了什么，这种规律都是自我存在着的。”

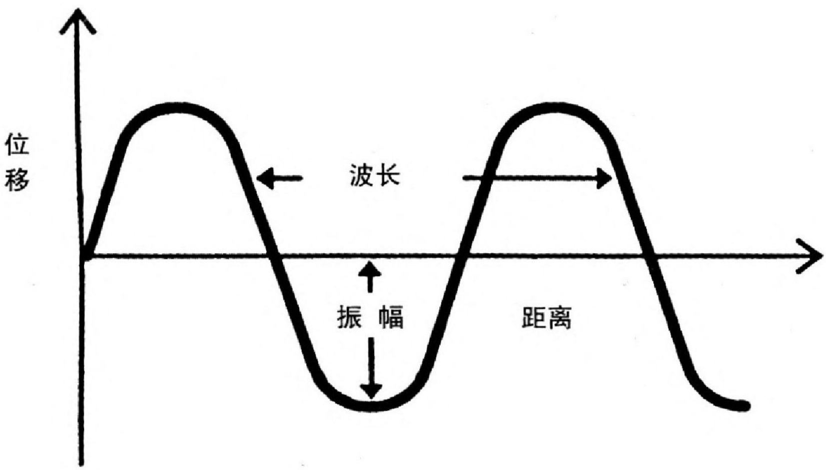



图3　波长的意义

在牛顿时代，神学中的宇宙模型把绝对空间和绝对时间定义为造物主的属性，是毫不困难的。但是，在牛顿物理学里，这样的绝对空间和绝对时间到哪里才可以找得到，这是远远不清楚的。牛顿本人主张，绝对空间（以及绝对时间）应与太阳系的中心重合；另一些人后来进一步认为，应该用所谓“固定恒星的参考系”作为绝对空间的框

架。这些遥远的恒星，由于距离上的原因，看起来是固定不动的。

实际上，在牛顿力学里，要承认绝对时间，就得否定绝对空间，否则逻辑上就讲不通。没有绝对静止、完全不动的状态——它完全决定于个人的观察角度，或者用术语来说，决定于观察者的参考系统。考虑一个最简单的例子，一个除了两个球以外，没有任何其他东西的宇宙。想象我们位于其中一个球上。如果两个球之间的距离在稳恒地增加，我们就不可能说，是这一个球还是那一个球在动，或是两个球都在动。在这两个参考系中，不可能有哪一个是有特殊地位的。这本放在桌子上的书，虽然对地球来讲是不动的，但对太阳来讲，就完全不是这么回事了。我们也不能假定太阳是固定不动的，或者宇宙中确有某个地方是绝对静止的。牛顿的理论，实际上否定了这个绝对空间的观念，许多科学家和哲学家都指出过这一点，其中包括牛顿同时代的人，例如伯克莱主教（Bishop Berkeley）和莱布尼兹，以及后来的庞加莱、马赫（Ernst Mach）和爱因斯坦。

正是因为宇宙中没有特殊的、延及各处的网格，也没有一个点，像一张图表中的原点那样，可以被唯一地指认出来，所以在所有的参考系中，牛顿定律描述的是同样的物理现象。我们看来是理所当然的：一位空中小姐在斟咖啡时，她会认为咖啡流动的规律，不论飞机是停在跑道上，还是在8000米高空以固定的速度飞行，都是一样的。我们只要考虑相对运动的作用就行了：当火车以每小时50千米的速度，从一位坐在月台上的铁路值班员身边驰过时，一位乘客向着火车运行方向，以每小时30千米的速度扔出一个苹果核。这样，苹果核相对于值班员的速度，就是每小时80千米。在数学上，这可以用一种叫做伽利略变换的方法来表示，这个方法讲的是，当两个或更多的观测者作匀速相对运动，也就是他们两两之间的相对速度保持不变时，如何把他们的测量结果联系起来。

尽管在逻辑上站不住脚，但绝对空间的观念在19世纪的物理学家的头脑中太根深蒂固了，于是他们很自然地设想，以太弥漫于整个绝

对空间，电磁波就是在这云雾状的以太中的扰动。以太是宇宙的属性，就像地图上的经纬一样。任何距离的测量，都可以相对于以太做出。但是对于这些物理学家来说，以太的含义，并不仅仅是一种宇宙的参考系而已。由于地球相对于绝对空间在运动，它就是在连续的“以太风”中疾驰。如果以太不是出于亚里士多德腐朽思想的一种心理上的神话，就应当找出证据，表明以太可以导致出某些实实在在的科学结果。著名的迈克耳孙-莫雷（Michelson-Morley）实验，其动机就是为了搜寻以太。关于这个实验和它的深远意义，我们将在下一章中再来谈。

## 牛顿物理学的预言能力

关于时间，牛顿方程式还有另一个令人惊异的特点：它们是“决定性的”。为了理解这个意思，我们可以想象任何一个系统，相互碰撞的台球也好，绕日运行的火星也好。牛顿的运动方程说，不管在观测的初始时刻位置和速度如何，也就是不管“初始条件”如何，系统的行为对过去和将来都是确定的。不论物体是受到电磁相互作用，像绕原子高速转动的电子那样，还是受到引力作用，像绕日运转的行星那样，牛顿力学原则上能使我们确定物体在整个过去和将来的行为，只要我们能够知道物体在某一个时刻的速度和位置。我们也许会想，将来是不确定的。但是按照牛顿的方程，将来是被详详细细地确定下来的。这种“决定论”是牛顿方程数学结构的一个直接推论。决定论与“因果律”密切有关，因果律说，每一个事件都有它的原因，而事件本身为其结果。在我们现在讨论的情况下，初始条件就是第一个原因，因为我们不问初始条件是如何导致的。

看上去，我们不得不下这样的结论：万事都是由宇宙的初始条件决定的——当上帝点燃大爆炸的导火线的时候，这初始条件就被确定了。哲学家们常常想“证明”源于偏见的信仰；他们列举出来的有利于自由意志的所有论点，都被牛顿的决定论击得粉碎。决定论也贬低上

帝和人类在宇宙演化中扮演的角色。牛顿的钟表机械式的宇宙，像是打在基督教神学心脏的致命一击。怪不得教会与科学的关系相当不融洽。

上一章提到爱因斯坦说过的一句话——过去、现在和将来之间的区别只不过是一种幻觉——这无疑是出自于牛顿力学的决定性的和因果性的结构，这结构也同样支承着广义相对论。牛顿力学是一个时间对称、决定性的理论，其中过去、现在和将来没有区分——这三者相互之间没有什么特别的关系。知道了某个任意时刻行星的坐标和速度，就可以完全确定地描述太阳系在所有“后来”和“以前”时刻的状态，只要把这些坐标和速度代入牛顿方程就行了。因此，对于一个力学系统，牛顿理论在任何一个时刻的描述，都在这同一时刻已包含其整个的过去和将来。时间倒转的对称性意味着“果”可以变成“因”，“因”可以变成“果”；这样一来，我们的“因果感”本身也就大有问题了。就牛顿方程而言，一场板球比赛的时间完全可以倒转，每一个球都完全可以回到投球手的手中。此时，“因”是球开始在草坪上先滚动后弹起时，球吸取的热量；“果”是球飞向球板，吸收声波，并弹回到向后退的投球手张开的手里。

## 庞加莱的回归论

牛顿的力学方程不具有内禀的时间箭头，于是没有理由说一个时间方向比其相反的方向更好。但是事情其实更糟。庞加莱提出的一个定理说，在一个足够长的时间间隔内，任何孤立体系（例如宇宙本身）将返回到它的初始状态；事实上，在一个无限长的时间中，它应当如此重复无限多次。这就是第一章中提到的斯多葛学派的时间循环。

庞加莱定理适用于一个大小有限的孤立体系：一个被限制在无摩擦的台面上运动的台球，迟早会以初始时的速度，回到它的初始位

置。这种永恒的反复，在台球游戏的例子中是容易想象的。但是，对于许多我们所感兴趣的体系，粒子（例如原子和分子）的数目是如此巨大，使得这“循环时间”比起宇宙的年龄（大约100亿年，也就是1后面10个零），要长出许多倍。即使如此，这种无止境的循环往复，还是暗中在摧毁时间箭头的基本观念，否定一切事物的演化，使演化的概念顶多只具有最肤浅的意义。庞加莱的循环论（或者叫做庞加莱回归论），不管它的缺点如何，成为了理论物理学家的麻醉剂，所引起的主要反应，大体上说来，就是坚持从主观立场上解释时间的不可逆转。

## 牛顿物理学的局限

牛顿物理学承认绝对时间的概念，但不承认绝对空间。如我们上面提到的，在向火车外扔果核的那一例中，牛顿方程的描述在伽利略变换下是不变的。这个变换讲的是，当两个观测者的参考系作匀速相对运动时，如何把这两个观测者对于同一个事件的记录联系起来。如果要求物理学与描述事件所用的参考系无关，那么这一不变原理是必需的。如果宇航员到了月球上，发现那里的物理定律不同，他们一定会感到非常惊异。因此，难怪不变原理在现代物理学中起着中心作用。但是，按照牛顿方程，两个观测者一方面由于相对运动而有不同的位置，另一方面他们对时间却有同样的感觉，而这种感觉是与参考系无关的。这就使得牛顿的或其他任何人的绝对时间观念得以继续存在。如果在早期某个时刻，两个观测者的表互相校准过，按照牛顿的世界观，两个人的表将永远显示同样的时间。这听起来非常合理，非常符合我们对时间的常识。

遗憾的是，常识往往是对真实世界的误导，这一点以后会很明显。其实，当我们注意到麦克斯韦电磁方程在伽利略变换下并非不变时，我们就意识到绝对时间是会有问题的。换句话说，电磁现象是随着伽利略参考系的速度而变化的。然而，伽利略变换之所以成立，正

是因为牛顿力学中没有一个绝对静止的状态。电磁波的不正常表现，暗示着存在一个特殊的参考系，也就是绝对空间幽灵的复活。奇特的作用看来是可能了，例如，光传播的规律似乎应该与观测者的相对速度有关。此后，洛伦兹发现了一个新的变换——洛伦兹变换——在此变换下麦克斯韦方程是不变的。这一变换和伽利略变换大为不同，特别是它把两个作匀速相对运动的观测者的空间和时间坐标混在一起，从而使得光传播的规律与速度无关。洛伦兹的这个形式古怪的变换，当时被认为只是一种雕虫小技，因而未受到重视。直到爱因斯坦的狭义相对论1905年问世，它的命运才改观。狭义相对论我们在第三章中将要谈到。

除掉这一点小困难以外，牛顿物理学对物体在引力和电磁力作用下的大尺度动力学行为，无论是从树上掉下来的苹果，还是绕日的行星轨道，都给出了很好的描述。然而，对于看不见的原子和分子的微观世界，也就是构成物质的基本单元，情况又是怎么样呢？

在我们的故事现在讲到的那个时期——即19世纪末——

还没有一个被普遍接受的物质原子论。但是有利于原子论的证据，是在不断地越积越多。做出主要贡献的，有玻耳兹曼，他奠定了气体动力学理论的分子论基础；有麦克斯韦，他把黏滞性与分子行为联系起来，玻耳兹曼把这比作“一首绝妙的交响诗”；此后，还有爱因斯坦，我们将在第四章中再谈到他的关于布朗运动的分子理论（1905）。所有这些人的工作，都是运用牛顿力学去描述分子的运动。

反对物质原子论的论调，主要是由后来被叫做“维也纳派”的“逻辑实证主义”的先驱者们激发起来的。这一哲学学说的倡导者，其中相当一些是有影响的人物，例如马赫和德国化学家奥斯特瓦尔德（Wilhelm Ostwald），他们坚持认为，任何有关所谓原子论的陈述都是没有意义的，因为我们没有办法直接证实原子和分子的存在。另

一方面，玻耳兹曼深信原子论是科学上必需的，因此他不断地与这些固执的反对者争论。逐渐地，这些反对者在论战中处于下风。

到了20世纪初期，原子论已经牢固地建立起来。当时大多数物理学家认为，现在对一切事物基本上已经完全了解了。也许只有很少的几个小问题，是需要解释的——例如，某些物质吸收热量的准确方式，以及某些原子蒸汽辐射出的奇怪谱线——但物理学家们声言，只要再过几年，这些问题就可以解决了。让我们仅仅举一个例子。迈克耳孙在1903年信心十足地说道：“所有比较重要的基本定律和物理科学的事实，都已经被发现，它们已经很稳固地成立，甚至连因为有新的发现而要对它们进行补充的这种可能性，都极其微小。”温伯格写道：“这段话以后一直被物理学家们当做笑料。”然而，人们仍然一次又一次地拿科学上的运气打赌。1928年，后来获诺贝尔奖的玻恩（Max　Born）发表评论说：“我们所了解的物理学，将于六个月内大功告成。”核力的发现，很快把这种想法送进了垃圾箱。但是即使到了1988年，霍金在他的畅销书《时间简史》中还说，他相信“可以谨慎乐观地说，我们对自然的终极规律的探索，现在也许接近了尾声”。让时间来检验他的预言吧。

无论如何，原子论的兴起，敲响了牛顿物理学在某些领域的丧钟。日渐增多的证据表明，牛顿物理学对于描述高速、极大质量和极小质量的情况，已经失效。下两章要叙述的两个革命，在20世纪的头25年中，冲进了牛顿的物理世界。量子力学使我们对“基本”微观世界的认识面目一新，爱因斯坦的相对论把绝对时间的观念一扫而光。正如斯夸尔爵士（Sir John Squire）模仿坡普的诗句写道的：“光明不再继续了，恶魔高声地在嚎叫；让爱因斯坦出世吧，使世界重现光明。”

## 时间失去了方向

作为我们日常生活中一个基本特点的时间，它的本质，在牛顿物理学中一直是含混不清的。时间是牛顿为了以数学方式描述运动的概念而引入的；他把运动定义为位置随时间的变化，而时间是一个基本量，本身没有定义：运动是用时间来解释的，而不是时间用运动来解释。我们可以把一个物体任意置于空间一处，但我们不能控制它在时间中的位置。

在一个牛顿宇宙中，相对做匀速运动的钟，不论它们的位置和速度如何，时间流逝的快慢都是一样的。一般人对这种时间观是最感惬意的。然而这种时间观被爱因斯坦的相对论否定了，这我们将在下一章中再谈。

牛顿的时间中有一个佯谬。人们的经验是，时间永远是向前走的。时间的流逝才使我们能够观察到运动，但是时间箭头的起因一直没有解释。牛顿的运动方程尽管具有很大的能力，但是它们产生的结果却是与直觉相反的：时间的对称性使这些方程对时间的方向漠不经心。经过庞加莱回归，这些方程确保历史会无限地重复。它们的决定性，如再加上关于一个系统的足够信息，就可以断言该系统所有将来和过去发生的事件。牛顿的这种自然观难怪是浪漫派诗人所憎恶的。济慈（Keats）在他的《拉米亚》一诗（1819）中写道：

科学将剪断安琪儿的双翅，

用规则和准线打破所有的秘密，

把幽灵赶出天空，把地精赶出地洞，

把天上的彩虹拆散，叫它们永远不再编织。

如果牛顿力学是普遍适用的，我们就不得不说，所有可能发生的过程，都可以用其组成部分的原子和分子的运动来表示。因为牛顿力学是决定性的，一个系统将来和过去的行为，就可以从该系统任何一

个时刻的信息得到断言。我们的大脑既然也是由原子和分子构成的，自由意志就不可能存在。法国哲学家波格森（Henri Bergson），和其他许多人一样，为这幅大有问题的世界图像感到忧虑，因为他觉得“经典物理学中，所有的事物都是一次性给定的：变化只是将然的否定，而时间也只是一个参数而已”。同样的困难也使得另一个法国科学哲学家科瑞（Alexandre Koyr）认为，把牛顿力学描述的运动看做是“一种与时间无关的运动，一种在‘非时’时间中进行的过程——这样的时间概念，就和‘没有变化的变化’同样地令人难解”。

牛顿力学造就了一个充满活力的理论机体，这机体，人们至今仍然把它应用在各式各样的场合：从台球的运动到星系的形成，到空间探测的技术。行星、导弹、火箭、卫星和诸如“旅行者号”那样的空间探测飞船，它们的轨道都是以这300岁高龄的理论作依据的。正如以色列（Werner Israel）和霍金所述：“它工作的精确性令人难以置信——对于地球的绕日运行，精度好于一亿分之一——而且它还继续在日常生活中发挥着作用。”对于描述行星轨道，以及其他只包含很少运动物体的“简单”系统，常常有这样的情况：它们的位置中微小的不确定性没有多大影响，原因是牛顿方程对非常相似的现状，总给出非常相似的将来。但是现实世界里要考虑的，往往是包含众多物体的复杂系统，而我们对它的信息又不可能掌握完备，在这种情况下，上面说的那种决定性和实际就很少有关系了。在绝大多数情况下，例如第一个太空人从空间看到的庞大的地球天气系统，对现状的描述中最微小的不确定性，也会导致完全两样的将来。失之毫厘，谬之千里，拉普拉斯梦想的决定论，因此被一笔勾销。他本来认为，只要我们知道了在某个任意时刻，宇宙中所有粒子的位置和速度，我们就可以对一切做出预言。随着决定论的灭亡，重新发现一个统一的自洽的时间观的可能性，终于来到了。

# 第三章　时间使爱因斯坦受挫

相对论教我们对时间要警惕。

——伦德勒（Wollgang Rindler）

《相对论精义》

阿尔伯特·爱因斯坦26岁的时候，就把有300年历史的绝对时间观念摧毁了。他推翻了牛顿物理学的整个基础，对现实进行了革命性的重新评价，赋予时间和空间全新的意义。这就是“相对论”。它给出了许多完全新颖的结果，其中有“时间膨胀”（即一个人的时间相对于另一个人变慢），时间经过空间“蛀洞”的行进，以及“新生儿自谋杀”的异乎寻常的景象，即时间旅行者回到出生时刻从而“谋杀”了他们自己。爱因斯坦粉碎了几乎所有关于时间的常识。但是，如我们将要看到的，时间问题中有一个基本方面被他避开了。他没有考虑时间箭头。

如果我们认为，一切事情都是在爱因斯坦的相对论论文发表的一夜之间改变的，那就错了。摧毁绝对时间观念的种子，早在17世纪就被丹麦人罗耶默（Ole Roemer）播下了。确实，按照爱丁顿的说法，我们现在所理解的时间，是被罗耶默发现的。当他1675年在巴黎天文台研究木星卫星的不规则运动时，首次指出光信号具有速度。他告诉一些科学院院士说，由于光线不是瞬时地而是逐渐地传播——这种看法持续至今，因此离木星最近的一个卫星下一次的掩食时刻，将比根据以往观测推算出的晚10分钟。

一旦人们开始去想光线从蜡烛或电灯泡传到眼睛需要时间而不是

瞬时地传播过来，就会搞清楚当我们注视天空的时候，我们看到的很远的恒星或星系是它们很久以前的样子。直到1728年，罗耶默的看法才被英国天文学家布拉德雷（James Bradley）所证实。这以后，光的有限传播速度的概念才被广泛接受（现代的光速值大约是每秒300000000米）。然而罗耶默的工作还标志着，最质朴的绝对时间的概念开始终结。绝对时间设想，我们看到的宇宙深处所有角落发生的事件，都和地球上的时间同步。爱因斯坦表明，即使考虑到在把这些事件的信息传过来时，光具有速度，绝对时间也不能够成立。处在不同运动状态的观测者，他们测量的时间不再彼此相同：相对于一个不动的观测者，一座钟所显示的时间决定于它的速度。如果还要考虑引力的话，那么这个时间还要与钟在空间中的位置有关。

公众对爱因斯坦的传统印象是一个白发苍苍、为人和善的古怪老头。但实际上，当他在第一次世界大战前的几年里震撼科学世界时，他是一个干净利索的年轻人，黑色卷发并留着胡须。一个传记家曾这样写过爱因斯坦："显而易见，他看来特别乐于与女士们交际。这种感觉常常是相互的。这个刚成名的年轻人，有一头浓密的、乌黑发亮的波浪式美发，一双大而明亮的眼睛和不拘礼节的风度，明显地具有迷惑力。"这也是一个身负重大使命的年轻人。爱因斯坦的梦想是创立一个与实际世界相符的描述，它不受人为偏见的影响，在它描述的世界里，客观存在是至高无上的。他抱负的志向是使物理学规律完全摆脱观测者的影响，为此他坚持认为，所有的观测者，不论他的位置和运动状态如何，必须同等地看待。任何个别的观测者或"参考系"都不具有特殊的地位——物理学规律必须与这些琐事无关。爱因斯坦远征的第一个阶段是狭义相对论，它建立在一个新的原理之上，即光的速度对所有的观测者都是相同的，不论他们自己的速度如何。后来这个理论扩展到对引力的解释，也就是广义相对论。

## 迈克耳孙-莫雷实验

在相对论问世之前，以太的观念就已经不得不离开历史舞台了。以太是由日常经验启发出来的。声音是空气中的波，涟漪是水中的波。所以19世纪的物理学家自然地想到，光同样必须是在某种东西中的振动，他们就把这种东西称为以太。正如派斯（Abraham Pais）所说，以太是“一个富有奇趣的假想介质，它的引入是为了解释光波的传播”。以太也为测量绝对空间提供了一个标度。它可以被想象为一张笼罩着整个宇宙的无形的网，就像地图上的经纬线一样，可以作为测量距离的参考系。

我们前面谈到过的麦克斯韦电磁理论，它的成功启发了一些人去做实验，目的是测出地球绕日运行穿过以太的速度——也就是“以太漂移”的速度。事实上，正是这其中的一个实验——1887年在克里夫兰的切斯应用力学学院做的——让以太的幽灵寿终正寝。这是搜寻以太的实验中最著名的一个，是由物理学教授迈克耳孙和他的同事莫雷——一位化学教授，一起完成的。基本上，他们的实验是重复迈克耳孙以前的一个实验，在那个实验里误差把结果搞得模糊不清。他们用一个光源向两个方向发射光束，两束光被与光源距离相等的两面镜子反射回来。其中一束光的方向，是沿着设想的地球相对以太运动的方向，另一束的方向与其垂直。

迈克耳孙和莫雷期待发现，在两个相互垂直的方向上光线传播的速度会有不同，这差别可能就是由于地球在以太中穿行的运动而导致的。比如，在地球轨道运动方向上传播的光，会相对以太风逆行。这样，这束光的速度就会比其他方向上的光速慢。分析光的运动就可以知道，垂直于以太风方向的光束，比沿着以太风方向的光束提早被反射回来。两束光被反射回到光源处的时间差，可以计算出来并与实验相比较。但是迈克耳孙和莫雷没有发现一点时间差。为了防止地球绕太阳转动时，以太风的方向发生改变而影响实验结果，他们在一年的不同时间重复同样的实验。但是不管他们如何努力，还是没有观察到时间差别。这就是说，根本就没有以太。

这是一个令人惊奇的结果。诺贝尔奖获得者密立根（Robert Millikan）当时认为，它是一个“不合道理的，看上去无法解释的实验事实”。看来，不论光是顺着以太风或地球运动方向走，还是逆着走，光的速度并没有改变。如果以太站不住脚了，力学体系本身可能就需要重新改写。“静止的以太”提供了一个绝对的参考系，它是牛顿心理上需要的，虽然牛顿力学也认为以太是多余的：因为没有办法去探测绝对运动——如果你在火车里放手让一个球落下，则无论火车是停着还是以不变的速度在行驶，球都将是垂直地下落。人们所能够观察到的，只是两个物体彼此间的相对运动，并不存在什么绝对的参考系。

但是，仍然有一种看来是从逻辑上对以太的需要，因为电磁学并不满足此种相对性原理。电磁作用随着观察者而异，并且看上去确与某个绝对参考系有关：当时就有不同的方法解释发电机如何把运动转变成电力，电动机又如何把电力转变成运动。因为以太的观念已经站不住脚了，爱因斯坦需要一个新的理论对所有的自然现象给出一个统一的描述。结果是他改写了物理学，让迈克耳孙一莫雷实验的否定结果成为新原理的自然后果。

许多作者声称，迈克耳孙一莫雷实验与另一些或早或晚的实验一起，是对以太致的悼词。这样说无疑是过于简单了。许多著名的物理学家，仍在努力使迈克耳孙一莫雷实验与以太的假说相符合。这当中最有名的要算是荷兰的洛伦兹和爱尔兰的菲兹哲罗（George Francis Fitzgerald）。他们试图利用上一章中提到过的洛伦兹变换，用物体穿过以太运动时的物理收缩来解释迈克耳孙-莫雷的实验结果。这样，以太的假说就可以仍然成立，不过要以一种未经解释的运动物体的畸变作为代价。我们将会看到，这种长度收缩，与爱因斯坦所揭示的世界中的效应相近，我们以后必须习惯于这种效应。

洛伦兹实际上已经接近了狭义相对论的公式，但是他不能摆脱牛顿的绝对时间“经典”观念的束缚，并且紧抱着以太理论不放。法国数

学家兼物理学家庞加莱，对牛顿力学造成的问题看得很清楚，他问道："以太究竟是什么，它的分子是如何排列的，它们是相互吸引还是相互排斥？"并且他热切期望着如爱因斯坦后来提出的根本解决办法。他说道："也许我们必须建立一种新的力学，对它我们只能够管中窥豹……在这个新力学中，光速是一个不可逾越的极限。"1904年庞加莱甚至于编造出一个"相对论原理"。但是按照爱因斯坦自己的说法，看来庞加莱至死都没有搞懂狭义相对论的物理含义。

爱因斯坦本人很久以后才知道上述物理学家的种种努力，他基本上是独立地得出他的理论的。他当时并不熟悉那些在物理学杂志上发表的、时新的研究论文的内容。确实，他一点不知道洛伦兹1895年以后的工作；特别是，如我们将会看到的那样，他从没有听说过洛伦兹变换，但这个变换却在他自己的研究结果中再现了。我们甚至都不清楚，爱因斯坦是否认为迈克耳孙一莫雷实验对他后来的狭义相对论起了决定性的影响，虽然1916年他的朋友心理学家沃斯默（Max Wertheimer）在柏林采访他时，他明确地说过是受到过它的影响。然而，在1954年的一封信中，他坚持说："在我自己的研究过程中，迈克耳孙的结果对我并没有多大影响。我甚至于都记不清楚，当我写关于这个题目的第一篇论文的时候（1905），我是否知道这一结果。在我个人的奋斗中，迈克耳孙的实验没有起过作用，或者至少是没有起过决定性的作用。"

## 通往狭义相对论之路

这样，我们现在就讲到爱因斯坦本人。他1879年3月14日上午11点30分出生的时候，他母亲鲍琳相当吃惊。这孩子的后脑很大而且棱角分明，她怕这孩子是个畸形儿。他发育得很慢，语言能力又非常差，周围的人担心他可能永远不会说话。当他8岁那年在慕尼黑上中学时，他的希腊语教员对他说，他将来不会有大出息。1894年他家搬到意大利，爱因斯坦被留了下来，在他不喜欢的学校里继续受煎熬，

这是因为这个学校的严格制度，以及德国军队需要征召16岁以上的青年。他很难让学校喜欢他，他“早熟，半盲目自信，几乎目空一切”。希腊语教员甚至建议他应当退学。确实，他父母走了不到半年，他就也跟着翻过了阿尔卑斯山。爱因斯坦后来写道：“我的班主任老师把我叫了去，要我退学，但不给我以后能保证我进大学的文凭。我说，‘我从来没有做过任何错事啊’，他却说，‘你只要一露面，班上就对我不尊敬了。’肯定地说，我自己是希望退学，跟着父母去意大利。但对我来说最主要的原因，还是那里呆板的、机械式的教学方法。”一旦离开学校，爱因斯坦高兴得像一只出笼的小鸟，在回到父母身边以前，他抓住机会在意大利北部作了长途旅行。

1895年爱因斯坦决定去试一下运气，报考苏黎世的联邦工艺学校（现在叫做ETH），希望以后成为一名电气工程师。但是他没有通过入学考试。在阿劳的一个瑞士州立学校补习了一段时间后，他才考上。在ETH学习期间，一个教过他的、俄国出生的老师闵可夫斯基（Hermann　Minkowski），有一次把他形容为一只“懒狗”，“在数学上一点都不用脑筋”。无疑，爱因斯坦是靠了他的朋友和同学格罗斯曼（Marcel　Gross-mann）的笔记，才补上了他没有去听的课。闵可夫斯基后来在爱因斯坦学说的发展过程中，起了关键性的作用。1900年爱因斯坦毕业后，在苏黎世做家庭教师，并且兼代课教员。到了1902年，由于格罗斯曼父亲的推荐，他在伯尔尼的瑞士专利局找到了一份差事。爱因斯坦申请的是二等技师，他却只得到三等技师的职位。就是在这看来不大可能有大作为的职位上，他创立了与牛顿理论同样宏大的科学理论。在专利局他遇到了贝索（Michelangelo Besso），他是一位工程师，后来成了爱因斯坦的终生挚友。在爱因斯坦的第一篇相对论的论文中，贝索是唯一被爱因斯坦致谢的人。

1905年是爱因斯坦的“奇迹年”。他是一个非正统的人，学术上有点小名气。他能在同时代人中出类拔萃，是因为他真正的天赋品质，使他能够从传统思想的长期束缚中解放出来。爱因斯坦把牛顿力学和

电磁学的种种不解之谜推到它们的源头，对整个物理学用全新的基本原理进行阐述。这是一个真正的革命性的跨越。从这些基本原理演绎出的结果表示，我们的有限经验所给出的时间和空间的“常识”，可能会欺骗我们。

著名的摄影家哈尔斯曼（Philippe Halsman），有一次把爱因斯坦不喜欢穿袜子的习惯与他的这一品质联系起来。他向这个大人物问到他的怪癖，爱因斯坦的秘书海伦插话说：“教授从来不穿袜子。即使是罗斯福总统请他去白宫，他也不穿袜子。”爱因斯坦解释说：“我发现大脚趾总爱把袜子顶穿一个洞。所以我再也不穿袜子了。”也许这可以从爱因斯坦1901年服兵役的记录中看出一点眉目，在“疾病或缺陷”的栏目之下，写着“平足和汗脚”。但是哈尔斯曼用了一个更浪漫的看法，“这一细节看来是爱因斯坦绝对独立思考的象征”。

## 绝对时间被废弃

在令人困惑的实验结果面前，科学家们面临选择——或者是让现有的理论七扭八歪，硬是去凑合实验结果（这相对来说比较容易，但常常无效），或者是创立他们自己的新理论（这比较难，甚至极难）。爱因斯坦勇敢地选择了后者，他把日常经验所给的印象完全抛到一边。从形象到抽象的转变，在现代物理学中一直继续着，并取得了高度的成功，这完全应了爱因斯坦所说的一句著名的话：“大自然扑朔迷离，但没有恶意。”虽然有涟漪在水中传播、声音在空气中传播这样的事实，爱因斯坦还是断定，电磁波是一种基本的实在，它并不需要经过以太传播就可以存在。他把从麦克斯韦电磁理论开始的物理学理论，纳入系统的抽象数学表示。对那些朴素的物理模型他不感兴趣，虽然它们给出形象的实在，让人感到舒服。一个典型的例子是当时流行的原子图像——葡萄干布丁模型，一个球状、带正电荷的“布丁”上面，点缀着带负电的“电流”。他感兴趣的是真理——不论它会给出多么奇怪和令人惊异的结果。

例如爱因斯坦认为，电磁现象的描述中有些地方不正常，无法让人接受，这就是当电和磁同时存在时的作用。回忆一下，当时人们对电动机和发电机的解释是，它们有不同的工作原理。但爱因斯坦确信，电动机把电转换为运动，发电机把运动转换为电力，它们所依据的物理是完全相同的。人们习惯上认为导体运动会产生与磁场运动不同的结果。爱因斯坦认为这种看法是不合逻辑的。它意味着应当可能探测到绝对运动。他没有去用一种“特殊作用”来强使这些疑难问题归顺现有的物理定律，而是用电动机和发电机中的相对运动，把对它们的描述统一起来。

1905年爱因斯坦提出了两条全新的物理学基本原理。它们出现在他的第一篇关于相对论的科学论文中，这篇论文发表在权威的德国物理杂志《物理学年刊》上，题目是“论运动物体的电动力学”。这两条原理涉及以不变速度运动的观测者，是狭义相对论的基础。它们是：

1.“相对性原理”：宇宙中所有各处的物理规律都是相同的，不论观测者的运动速度如何。

2.光速是一个常数，它与光源的运动无关。

爱因斯坦的第二条原理，即光速的不变性，似乎有些耸人听闻。这就相当于，当你测量一颗步枪子弹的速度时，无论这子弹是一个没有移动的步兵向你射出的，还是由一架飞行中的超音速飞机向你发射的，你发现它们的速度一样。但实际上子弹的速度并非如此。那么为什么偏偏光速就应当是不变的呢？爱因斯坦证明，无论两个观测者之间的相对运动是多快，在他们各自的参考系中测到的光速都是相同的。这样，爱因斯坦给了常识狠狠的一击。爱因斯坦宣称的光速不变性，以及相对性原理对所有的物理现象是普适的，把当时流行的说法一下颠倒过来了。这样他的思想就超过了牛顿，牛顿理论只涉及纯力学现象，而他的理论涉及整个物理学——这的确是一步勇敢的跨越。

蕴含在牛顿运动三定律中的牛顿力学，也同样是相对论性的：宇宙中没有一个特殊的参考系，其他参考系与之比较可以得到一些绝对的量。正如我们在上一章中看到的，如果宇宙中只有两个物体，并且它们之间的距离在不断增加，我们不可能判断是这一个还是那一个物体在动，还是两者都在动。相对运动的思想可以上溯到伽利略，他研究了当船静止以及船以不变速度航行时，船上的蠓虫、苍蝇、小飞虫和鱼的运动情况。他写道：“当你仔细观察这些东西时……让船以任何你喜欢的速度前进，只要它的运动是平稳的、没有任何摆动，你将不会发现这些东西的运动中有一点变化，你也不会从它们中任何一个的运动中发现船是在航行还是静止不动。”但是牛顿（尽管如此，他还是喜欢相信有一种绝对静止的状态）同时还暗含地用了另一条基本假设，也就是关于绝对时间的假设，它在整个宇宙中都是一样的。在牛顿物理学里，时间是以同样的快慢流逝的，不管观测者的速度或位置怎样。

因为对所有以恒定的相对速度运动的观测者，牛顿力学定律都成立——这就定义了“惯性参考系”——不同参考系的时间和空间坐标就通过伽利略变换联系了起来。用伽利略变换，可以把对运动的“看法”从一个参考系变换到另一个，比如说从伽利略船上的一只蠓虫，变换到站在岸上的一个水手。有重要意义的是，这个变换不能用于主导光的行为的电磁学定律。如果我们代之以爱因斯坦的第二条基本假设，使光速以及物理学规律不管观测者的速度如何，都保持不变，则这样的变换称为洛伦兹变换，它保证了描述光和其他现象的麦克斯韦方程，不论观测者的情况如何，都不发生变化。

在爱因斯坦以前，拉莫尔爵士（Sir Joseph Larmor）和洛伦兹也得出了这个数学变换。但是，在一次演示他的学说的威力时，爱因斯坦令人吃惊地从他的“第一原理”推导出了洛伦兹变换——仅仅根据他自己的基本原理，而不用参考上面两人的工作。事实上，他1905年6月发表的论文，并没有引用任何一篇参考文献。除此之外，他在论文

里只用了一句话，就把有两个世纪历史的以太的观念打发掉了。以同样速度在整个宇宙中流逝的绝对时间的概念，他也只是简单地将之摒弃。

绝对时间的被摒弃是一个意义深远的结果，它的重要意义在于检验第二条基本假设——光速的不变性，因为正是光速的不变性导致了这个结果。除了迈克耳孙-莫雷实验以外，对光速不因光源速度而变的另一件有利事实是1913年荷兰天文学家德西特（Willem de Sitter）对双星绕它们的公共中心转动时所发出的光的分析。然而直接的观测证实直到1963年才给出。这并不是说直到这以前，人们仍然在怀疑狭义相对论的正确性。对于许许多多其他的实验观察和预言，相对论的显著成就，已经足够建立起压倒牛顿理论的优势。理论和实验之间常常发生冲突，这自然是科学有别于哲学的显著之处。

## 接近光速运动时的新世界

尽管爱因斯坦对时间作了重新评价，牛顿学说的大部分，经过300年的考验仍然卓有成效。所以，一位宇航员1968年在第一次绕月航行返回途中，说道："我想，现在主要是伊萨克·牛顿在驾驶飞船了。"这句话突出表明，当年阿波罗计划是如何依赖于牛顿定律来计算空间飞船的轨道的。只有当物体运动的速度接近光速时，牛顿定律才会失效。这种高速运动的情况，与我们的日常经验常常迥然不同，除非是涉及光和电磁作用的场合（这也就是为什么麦克斯韦理论在牛顿的理论框架中非常别扭）。的确，在运动物体的速度远小于光速的极限情况下，例如一辆汽车在公路上行驶时，可以证明，狭义相对论的洛伦兹变换，就等价于经典物理的伽利略变换。换句话说，在这些情况下，狭义相对论就还原成牛顿物理或经典物理。因此，我们在这一章中要描述的、由相对论而引起的许多奇怪现象，只有当相对运动的速度趋近于光速时，才有重要意义。

长度收缩是一个很好的例证。爱因斯坦指出，接近光速运动的物体，在一个静止的观测者看来，会在运动的方向上变扁。这纯粹是一个相对论性效应：物体实际上一点都没有收缩，仅仅是观测者看来它变扁了。为了说明这一情况，我们想象有一列高速火车，即相对论快车。它只有一节客车，沿铁路线以不变的速度，相对于坐在站台上的观测者飞驰。当它的速度很大时，观测者会看到火车缩短了。但是从客车上的旅客看来，是站台在运动，所以是站台而不是火车，看起来变短了。收缩的程度决定于运动物体的实际速度：当这相对速度趋近于光速时，长度就收缩为零。物体在以相对论性速度运动时的现象，本身就是一个很有意思的课题：直杆变弯曲，自行车轮看起来像曲形飞标，等等。

如果站台上的人能够测量到的话，他们会发现，相对论快车的质量在高速情况下也改变了。与此同时，车上的旅客也会发觉，站台的质量变化了。这是因为相对论预言，运动物体的质量会增加。一个物体的“固有”质量，是指在相对它静止的参考系中所测得的质量。但是在另一个作匀速相对运动的观测者看来，该物体的质量随着物体的速度增加，质量增加的“洛伦兹因子”和长度收缩的因子完全相同。全世界的粒子加速器大量产生出的微粒子，它们这种相对论性的质量增加已被实验观测到，而且实验结果在定量上与爱因斯坦的预言完全相符。当物体的速度趋近于光速时，它的质量变成无限大，这样，就需要有一个无限大的力，把它加速到光速。这样一来，我们就可以明白，为什么不可能使一个有质量的物体达到光速，要它超过光速自然就更不用谈了。只有静止质量是零的粒子，才能以光速运动：光子就是一例——量子理论中的光子是跟电磁场联系在一起的。光子只能以光速运动。

## 同时性和时间膨胀

从“常识”的观点，狭义相对论最显著的特点来自于时间的相对论

化。同时的概念——事件在同一个时刻发生——决定于观测者的相对速度，而不是像牛顿认为的那样，是一种绝对的概念。如我们前面说过的，即使是在光速有限的牛顿世界里，也不可能看到世界在正好“现在”的样子，因为光线不是瞬时间到达我们的眼睛中，而是以光速传过来的。当你看表的时候，你看到的是有一点点“过时”了的时间，因为光线从表盘到达你眼睛中的视网膜，以及光信号被神经脉冲传送到大脑，都需要一定的时间。依照牛顿物理学，只要把时间作适当的改正，仍然能够重新建立起绝对时间，用以记录事件的发生，而且所有的观测者都可以接受这个绝对时间。但是爱因斯坦的相对论却不允许这样做。

为了说明“现在”所遇到的困难——也就是同时性是相对的——让我们再回到相对论快车。在车厢里面看来，从车厢正中的一盏灯发出的闪光，是同时到达车厢两头的旅客的。但是，站台上的观测者看到的是，闪光先照到车厢尾端的旅客，而后照到前端的旅客，因为火车的运动使车厢尾端迎着光线走（因而使光跑的距离短），而车厢前端顺着光线走（所以光线要多跑一段路才能追上它）。从站台上看，两束光到达车厢两端明显地不是同时的。如果车厢两端的旅客都有钟，则他们所分别记下的、光线到达车厢两端的时刻完全相同。但是，这样记下的时间，和用站台上的钟记下的时间相比，是有差别的。

同样令人惊奇的，是狭义相对论中所谓的时间膨胀现象。运动的钟要比静止的钟走得慢。假定火车里面的旅客和站台上的观测者用的是一模一样的钟。当火车停在站台上的时候，车上和车下都把钟对好，使大家的秒针每秒钟同时“滴答”一下（这就叫做“固有时间”，它表示用相对观测者静止的钟记录的时间，或者用另一个等效的说法，钟在自己的静止系中所记录的时间）。当火车运动后，在站台上的观测者看来，火车上的钟每两次“滴答”的时间间隔要比一秒钟长，变长的程度决定于火车的速度（再次由洛伦兹因子给出）。当火车的速度趋近于光速时，车上的钟两次“滴答”的时间间隔就增加到无限长。像

长度收缩一样，这也纯粹是一个相对论性效应。在车厢里的旅客看来，站台上的钟也按同样的程度变慢了。

时间的膨胀已经在实验上验证了许多次。有一种基本粒子叫做缪介子，它生成于地球大气10千米的高处，是由于极高速的宇宙线粒子的碰撞所产生的。缪介子的放射性衰变进行得很快（在它们自己的静止系中），如果不是由于它们的衰变时间在我们的参考系中膨胀了，它们的大多数决不会到达地面。因为用它们的静止系的钟测到的寿命计算，缪介子在衰变前只能走600米远。而用我们实验室的钟来计算，它们的寿命要延长8倍。物理学家们利用粒子加速器进行的许多类似的实验，同样证实了，缪介子的寿命可以由于加速到很高速度而大大延长。

## 双生子佯谬

狭义相对论中有关时间的另一件奇异的事，是所谓双生子佯谬，它是狭义相对论所有的佯谬中最早的一个。想象有一对长得一模一样的孪生兄弟达姆和迪姆，达姆出发进行一次相对论性（高速）的空间往返旅行，而迪姆留在地球上的家里。考虑离开地球的那段旅程，并且假定他们每个人都有一只特别的钟，这钟可以像灯塔那样，每隔5分钟发出一个脉冲信号。当达姆的速度增加时，迪姆在地球上收到的脉冲的时间间隔逐渐拉大；这就是说，从迪姆的观点看，达姆的钟走得慢了。这样，当达姆旅行结束回到地球时，他比迪姆要年轻（虽然他们的年龄都大了一些）。然而从达姆的观点来看，出现的情况应当是相反的——即他们重逢时迪姆应该比达姆年轻。当达姆开始他的旅程时，在他看来将是迪姆的脉冲间隔变大了，这表示迪姆的钟变慢了。

显然，他们两者的结果不会都是对的：两个人再相遇时，不会都说对方比自己年轻。对这个佯谬的解答，是我们要认识到，达姆和迪

姆并不是有从头到尾完全相同的经历。和迪姆不同，达姆在离开地球时，必须有一个初始的加速度，然后减速，再加速返回。最后，他应当减速，这样才能回到迪姆的参考系里。因为达姆不是以不变的速度旅行的（或者等效地说，他不是保持在惯性参考系中），我们就不能把狭义相对论的分析用于描述他所看到的情况，特别是用于他对时间的感受。这样，虽然这个“佯谬”不存在了，这个故事还是说明，的确发生了件很奇怪的事。在某种意义上我们可以说，多亏相对论，达姆才得以完成时间旅行，跑到迪姆的未来中去。我们只要把一只精确的钟放置在一架客机上，当它返程飞回后，把它所显示的时间和机场上的原子钟相比较，就可以实际观测到这个效应。然而，正像霍金所指出的，“用这种办法，飞行来回的次数要多得惊人，才能使人的生命延长一天。”

## 时空

狭义相对论的这些离奇古怪的效应，使我们对于时间的思考方式焕然一新。相对论学家喜欢说“时空”，这个概念把相对论的数学变得比较简单。它是出自于洛伦兹变换的数学性质，这个性质意味着空间和时间不应当单独处理，而是应当作为一个不可分割的整体来处理。这种空间和时间的融合首先是闵可夫斯基注意到的，在他学生爱因斯坦的狭义相对论的启发下，1908年9月他在科伦说：“从今以后，单独的空间和单独的时间注定要消失为阴影，而唯一继续存在的是两者的融合体。”

在狭义相对论里，时空的本质可以通过它的度规结构来理解，度规是一个抽象的、然而是基本的概念，它是与宇宙的几何结构联系在一起的。这种度规结构是内禀的，与任何观测者无关，这样的性质满足相对论的需求，可以确保物理定律的成立与速度和位置无关。在相对论中，“几何”性质——例如光线所显示的路径——是用时间和空间共同表示的，它们结合在一起，不可分离。所以，说时间和空间只是

单一时空的两个方面，这句话并没有什么了不起。虽然爱因斯坦在某种意义上排除了绝对空间和绝对时间的概念，但他却引进了绝对时空。然而，他把时空仅仅看做是所有事件的联络，这样一个人在时空图中，就像一条四维的“蠕虫”，它的每一张三维切片，就相应于这个人处在一个特定的时刻。在相对论里，我们可以把空间和时间作为四维存在来处理：空间是三维的，时间是一维的。但是从物理上说，它们是相当有区别的，这一点我们决不应该忘记。最重要的是，我们经历的时间——正如这本书要阐述的主题那样——是单向的，但是空间却没有这样的限制。爱丁顿指出过：“时间的最大特点是向前走。但物理学家有时常常容易忘记这一点……”我们再次回到了时间箭头——并且面对着狭义相对论的一个严重缺陷。

我们注意到，在双生子佯谬中，两个人的年龄都被认为是增长的。但是年龄增长的概念，作为一种时间的单向过程，在狭义相对论中并没有解释。这是因为，狭义相对论像它以前的经典力学一样，并没有区别时间可能经历的两个方向，也就是向前和向后。它只是说时间是一维的，并没有说时间是单向的。就相对论的时间对称的结构来说，也可以得出这样的结论，即留在地球上的迪姆，要比达姆年轻。但是我们必须认为这个结论是荒谬的，因为我们知道，事实上所有的生物，年龄都是在增长而不是在变小。但是，相对论本身，并没有解释为什么应当是这样。

## 加速度和绝对空间

狭义相对论还有另外一个较大的缺陷。它同样是与时间的作用有关，出现在有加速度的情况下（加速度是物体的速度随着时间的变化）。如我们已经反复说过的，由于牛顿力学中运动的相对性，绝对空间已经失去意义。可是对于加速度来说，情况就不同了。加速度是由于某种力例如引力所引起的，在牛顿理论中，加速度是绝对的。用另一种方式来说，无论观测者的运动状态如何，加速度总是相同的。

一个骑在马上的物理学家，可以争辩说是马在运动，或者是脚下的大地在运动。但是当他的坐骑急停，而把他从马鞍上摔下来，对这个过程他就没有异议了，因为无论从哪个参考系来看，例如从马、地面或行驶中的火车中看来，力和作为其结果的加速度都是一样的。但是对于本身就在加速运动的观测者，他们的看法与以不变速度运动的观测者是不同的。因此存在有地位特殊的参考系，即使认为地球向着物理学家加速，与认为物理学家向着地球加速是同样地可行。狭义相对论也只是对作匀速相对运动的观测者成立，因此也有这个问题。狭义相对论本身并没有给出任何解释，为什么必须赋予这些观测者以特殊的地位。他们只是直截了当地被放进基本假设里了。

爱因斯坦当时很清楚，引力破坏了他直觉上很吸引人的准则：物理定律应当与观测者的运动状态无关。换句话说，在狭义相对论对宇宙的描述中，把一个观测者的看法转换成另一个观测者的看法的数学操作——洛伦兹变换——不能用到引力上面。这问题的根源在于，与狭义相对论所暗指的相反，加速度是相对的而不是绝对的。可以用一个例子来说明这一点，同时也说明为什么包含有加速度的理论描述了引力。设想置于发射台上的空间飞船里有一位宇航员，他有一台放在洗澡间用的磅秤。如果他站在这磅秤上，磅秤将显示出他的体重。当起飞按钮按下，飞船加速飞出地球时，这位宇航员将感到自己重了许多，磅秤也显示他的体重大大增加了。假设发射失败，飞船骤然朝着地球掉下来。在这几秒钟内，这位倒霉的宇航员将在飞船里面自由飘浮，直到飞船撞上地面。在往下落的过程中，如果他把口袋里的钥匙掏出来，然后松开手，钥匙不会落到飞船的地板上。他也将会失重：在他的参考系里，是没有引力的。但是对于一个看着飞船往下掉的观测者说来，引力是再明显不过了。这样，加速度确实是相对的。爱因斯坦也受到过自由落体的启发——不过不是随飞船一起的宇航员，而是一个从屋顶上掉下来的人，这事发生在柏林。这个人侥幸没有摔伤，事后他告诉爱因斯坦，他没有感觉到引力的作用。

总之，爱因斯坦对于狭义相对论的局限性是很清醒的，并且从美学的基点出发，希望物理学彻底摆脱仍在苟延的特殊参考系。他开始着手他的广义相对论，去解决这更困难的问题，即给出物理学更概括的系统描述，这种描述对所有的观测者都适用，不论他们的相对运动状态如何。难怪，只有用比狭义相对论复杂得多的时间和空间关系，才能得到这种描述。确实，广义相对论的推导需要应用陌生的数学工具——张量计算。爱因斯坦为此费了多年心血，直到1915年才完成他的论文并准备发表。正如他在这期间所说："每走一步都是极其困难的。"

当广义相对论最后完成的时候，这个理论同时给出了一个漂亮的而且相当完美的引力理论。如果我们重新考察飞船里的宇航员，就可以知道为什么他不能说出他经受的是引力还是加速度。当飞船在空间中加速的时候，他也不能确定，他的磅秤显示的是引力的作用还是他自己的惯性——物体反抗运动变化的一种性质。爱因斯坦认识到了这一点，这使他在1907年提出了一条新的基本原理——"等效原理"。他强调说，这条原理适用于整个物理学。实际上，这条原理断言引力和加速度是等效的。

等效原理至少有两种说法。其一，即"弱"等效原理，可以回溯到伽利略和他的比萨斜塔实验，这个实验在传奇文学中被描写得很生动。伽利略发现，所有物体都以同样的加速度朝地球下落（在忽略空气阻力的情况下）。等效原理表明，用相对论的观点，我们也可以说是地球在加速向上而物体保持静止，这样，显然所有的物体的加速度就必须保持相同了。这个一直颇为神秘的难题，直到1914年爱因斯坦发表了他的论文，才被解释清楚。这篇论文说明，一个均匀的引力场完全等效于一个适当的加速度，对在任何实验室进行的实验，结果都是如此。等效原理还说明，狭义相对论是一个纯粹局部的理论：没有一个实际的观测者不在经受加速度，因为我们的宇宙是被引力统治着，宇宙物质以恒星、行星等形式散布在整个宇宙里面。等效原理的

第二种说法，即“强”等效原理，是爱因斯坦主张的，它认为所有的物理规律，对于宇宙中任何地方、任何时刻的所有观测者都是相同的，不管运动的情况和引力如何。对爱因斯坦来说，这个原理使他能够离开狭义相对论，而进入一个关于宇宙的理论，这一理论必须超越基于狭义相对论的局部描述。

## 通往广义相对论之路

为了创立广义相对论，爱因斯坦经历了8年艰辛、专心致志的努力。在此期间，新见解不断在他头脑中闪现，同时他也一次又一次地走进死胡同。直到最后，一个崭新的、闪耀着智慧光辉的理论终于出现了。1909年7月初，爱因斯坦辞去了他在伯尔尼专利局的工作，去苏黎世大学做教授。这是他在其后5年中得到的几个教授职位的第一个。

从1907年年底到1911年年中，爱因斯坦对有关引力的课题一直保持沉默。虽然他仍然在花费很多时间思考这个问题，但是刚刚诞生的量子理论（我们将在下一章讨论这个理论）使他也用去了不少时间。虽然爱因斯坦已经对这个新诞生的理论做出了重大的贡献，但在这个时期，它仍然占据着他的心思，并且在他此后的一生中，一直是使他忧虑的主要原因。在这个期间爱因斯坦的工作几经变动，1912年8月，他从布拉格的卡尔·费尔迪南德大学回到苏黎世，这看来对广义相对论的数学发展有决定性的意义。

当他离开布拉格的时候，他已经确信，时间和光线的轨迹都要被引力弯曲。但是这个想法必须要有坚实可靠的基础。刚刚回到苏黎世，他就转而向他的老友和同学格罗斯曼求助，这时候格罗斯曼已经是几何学教授和ETH的数理学部主任。他对格罗斯曼说：“你必须帮助我，不然我就会疯了！”

# 平直的和弯曲的空间

为了了解爱因斯坦是如何解决引力问题的，我们首先必须考虑一下我们日常所经验的世界的几何。公元前320年到前260年生活在亚历山大的古希腊数学家欧几里得，对此几何有过详尽的阐述。爱因斯坦发现，欧几里得几何（欧氏几何）只适用于空间中某些限定的区域。由度规结构描述的那些几何性质，在地球上是非常有用的，但是应用到宇宙的大尺度结构上就不行了。

考虑时空最简单的办法是把时空当做只是空间，同时用光的速度作为一个量杆（请记住，光速是绝对的）。一段时间的间隔可以转换成一段空间长度，只要简单地用光在这段时间内走的距离来表示就行了。天文学家们常采用光年来表示星系以及星系之间的距离，1光年大约是10万亿千米，同时也常用另一个叫做秒差距的单位，它等于3.26光年。这样做是为了避免太多的零出现在距离的表示中。例如，采用这样的距离单位后，太阳的距离仅仅是8光分（光在8分钟内走过的距离），天狼星的距离是2.7个秒差距，双子座星系团的距离是3亿5000万秒差距。

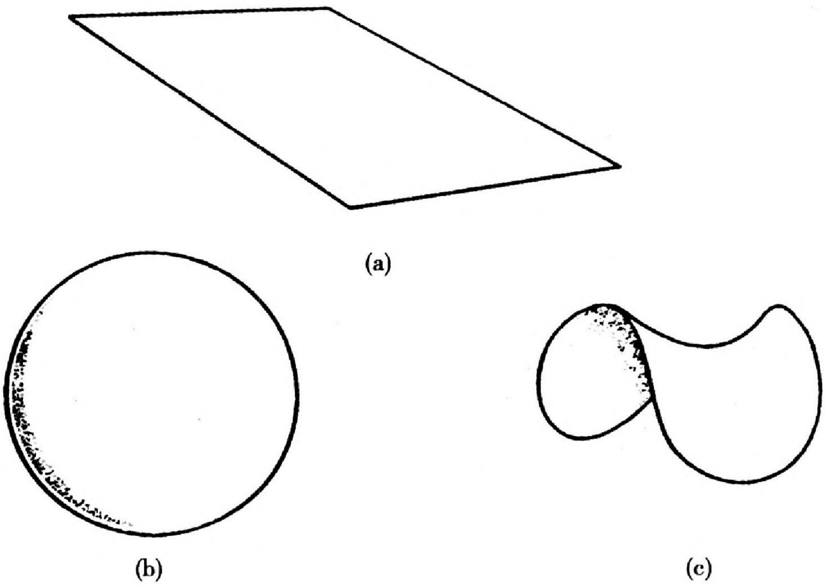



图4　平直的（a）和弯曲的[（b）和（c）]空间。球面（b）的曲率是正的，而鞍形面（c）的曲率是负的。

在狭义相对论中的度规性质意味着时空几何是平直的，像一张铺着绿色厚毛呢的台球桌面那样。但是在广义相对论中，我们必须熟悉弯曲时空的概念。从直觉上，每一个人都知道一个平面，即一个两维空间，是什么意思。一张平展地放置在桌面上的纸，就表示一个平直空间（它没有曲率）。而另一方面，球面却是弯曲的。这些两维空间或者表面（它们在数学上叫做流形）很容易阐明，因为它们嵌在我们非常熟悉的三维空间之中。我们不大可能直观地想象，高于三维的几何结构是什么样子，除非在某些神秘的感受下或许可能。然而非常重要的一点是，我们要认识到，一个空间的平直或弯曲，完全是这个空间的内禀性质，并不需要一个更高维的空间作为参考对照物。

平直表面的几何与弯曲表面的不同，这一点具有基本的意义。孩子们在学校学的是平直空间的几何，它在两千多年以前就被欧几里得详细阐明了。每一个中学生都知道，三角形的三个角之和是180°，以及半径是R的圆的周长是2πR。爱因斯坦这样讲到过：“欧几里得几

何……是一座宏伟壮丽的大厦，在它高耸的阶梯上，你会被认真尽责的老师们紧追不放，为它花费掉无数个钟点。”但是实际上，它的结果只有对于平直空间才是正确的。画在一个球面上的三角形，它的三个角的和要比平面情况下的大，而球面上的一个圆的周长，要小于画在平面上的圆周长，具体的结果取决于球面的曲率。虽然我们不可能想象一个弯曲的三维空间，然而我们可以用同样的方法去推断它的存在。让我们来看一下所谓的“平面世界”，它是维多利亚时代的一位教师阿伯特（Edwin Abbott）1884年首先描述的。阿伯特讲述了一种叫做扁方先生的生物的奇遇，这种生物具有两维结构，没有上和下的感觉，只能保持在一个表面上运动。为了我们的讨论，让我们想象扁方先生处在一个球面上。它会很快发现它是生活在一个弯曲的空间中，虽然这在第三维看来是很明显的。为此，它只需要出发沿一条直线向前，然后在某一地点它就会发现，它已经回到了出发时的位置。实际上，确切说来，这个特点是扁方先生所居住的世界所具有的、整体拓扑或者大尺度形状的一个性质，而不是一个局部的性质。但是，扁方先生和生活在三维空间的我们自己，只需要测量这样的（局部的）性质，比如像圆的周长，就可以知道，这性质是符合欧氏几何的定律（这样我们就是生活在一个局部平直的空间），还是与欧氏几何不符（这样我们的空间就是弯曲的）。19世纪伟大的德国数学家兼天文学家高斯（Carl Friedrich Gauss，1777～1855），认识到了这一点并且做了许多实验，去探测我们的三维空间偏离平直的程度。但是，无论是他本人，还是后来继续做这件事的人，都没有在地面实验中探查出空间的任何弯曲。这当然不会使我们感到惊奇，因为欧氏几何对我们来说是相当准确地成立，否则学校里就不会开这门课了。

然而，纯数学家通常是不考虑真实的物理世界的。在19世纪，他们开始构想任意维数和曲率的抽象空间，并且极为详尽地描述它们的几何性质。这个工作首先是高斯开创的，他的学生黎曼（Georg Friedrich Bernhard Riemann，1826～1866）发展了它，后来使这一理论臻于完善的，主要是克里斯多夫（Bruno Christoffel），李奇-卡

拉斯特罗（Ricci-Curastro）和李微-西威塔（Tullio Levi-Civita）。这些卓越的数学家阐明，度规结构可以告诉我们空间的情况，特别是它是平直的（欧氏的）还是弯曲的（非欧氏的）。

当这些发现刚刚被得出的时候，它们仅仅是使一个小圈子里的数学家从学术上感到兴趣的东西。直到爱因斯坦的工作问世以后，人们才广泛地认识到这些智慧之果所具有的深刻物理意义。除此之外，也只是由于爱因斯坦和他后继者的工作，时间才同样被纳入几何之中。如我们前面提到过的，闵可夫斯基关于狭义相对论的研究表明，为了数学物理上的目的，可以把时间作为像另一维空间那样处理。这样一来，不仅可以谈论平直的和弯曲的空间，而且可以谈论平直的和弯曲的时空。

刚到苏黎世的时候，爱因斯坦并不知道黎曼的工作，以及这件工作对于他本人正在思考的问题的重要意义。但是当他跟格罗斯曼讨论引力问题的时候，格罗斯曼告诉他说，他要寻找的东西是一种时空，它具有所谓的黎曼几何结构，这种结构完全不同于狭义相对论的欧几里得性质。

时空的关键特点是，即使它在大尺度上弯曲，在小尺度上也可以看做是平直的，正像一个人站在板球场上，会觉得地球看上去很平坦一样。这样一来，对于描述发生在时空局部区域的事件，狭义相对论和洛伦兹变换仍然可以成立。但是当这个区域扩展到时空曲率变得显著的时候，情况就不再是这样的了。这就像是，板球场在板球队员看来是平坦的，而它所在的那块大陆，在一个宇航员看来却是弯曲的。球面的半径越大，它的曲率越小，而且在任何一点的周围，看来是局部平坦的区域也就越大。

从欧氏几何转变为黎曼几何，这是使爱因斯坦得出他的后牛顿引力表示式的关键。起初他还得到了格罗斯曼的合作。1914年爱因斯坦迁居柏林，在那里他最后完成了广义相对论，他的这一论文题目

是“引力的场方程”，于1915年11月25日提交给普鲁士科学院。

## 广义相对论

时空弯曲的程度，是由宇宙中物质的分布所决定的：一个区域内的物质密度越大，时空的曲率也就越大。这样太阳附近的时空就要比地球附近弯曲得厉害，因为太阳的质量要大得多。广义相对论的宇宙中，引力已不再像以前我们理解的那样是一种力，它已经被转化到时空的几何（曲率）中去了。用爱因斯坦的新观点来看，可以说，引力产生于从狭义相对论的平直空间到广义相对论的弯曲空间的转换之中。

这样，我们对一些日常事件的看法，例如像对苹果落地这样的事件，就从根本上改变了。与其把引力想象成为某种神秘的力，经过空间作用在一段距离上，倒不如设想，像地球这样的大质量物体，使空间和时间发生了畸变。为了对这个说法有一个直观的了解，一个简单的办法，是把时空想象成一张平展的橡胶软垫。大质量的物体放上去，会使橡胶垫发生局部变形，变形的程度决定于物体的质量。太阳在我们太阳系中，质量远大于其他任何行星，所以它使时空畸变得最厉害。行星可以用大小不等的球来代表，这些球在橡胶垫上围绕太阳滚动，球滚动的路径也就是行星的轨道，它们都位于太阳附近的深“阱”之中。从树上掉下来的苹果，不是被一个力拉向地球，而只不过是滚进地球所造成的局部时空的“阱”里面罢了。

物体在弯曲时空中的运动规律，一般是不同于平直时空中的规律的。一个不受引力作用的物体，在三维空间中是做匀速直线运动的。而在有引力的情况下，新的规律则是物体沿“测地线”运动。测地线基本上就是在弯曲的或平直的时空中连接任意两点的最短的路线，只要这两点充分接近（图5）。在速度非常小、物质密度也非常低的情况下，测地线运动就退化成牛顿描述的运动。显然，广义相对论的这

种“退化”一定会发生，因为牛顿物理学所作的预言，在它所适用的范围内是十分成功的，这我们在上一章中已经讲到过。然而，对于牛顿无法回答的一些问题，爱因斯坦却可以用测地线运动来解释。

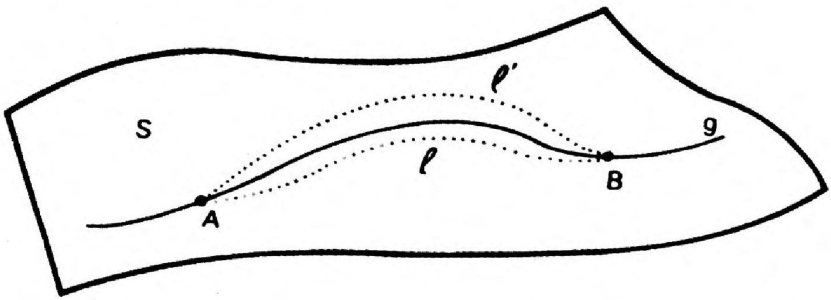



图5　测地线规定了广义相对论弯曲时空（S）中的运动路径。如果A和B是S中测地线g上足够接近的两点。则A和B质检所有其他的连线（l和l'）都比测地线长。在广义相对论中，地球围绕太阳的椭圆轨道被解释为，由于太阳的质量所造成的弯曲时空中的测地线运动。[录自W·仑德勒（W.Rindeier）《相对论精义》第106页。]

第一个例子是有关水星——它是离太阳最近的行星——轨道的一个很小但很重要的细节。虽然爱因斯坦在推导相对论的时候，几乎没有考虑到这个问题，但它却成了对他的新理论的一次辉煌验证。按照牛顿力学，一个单独绕太阳运转的行星，它的轨道应当是一个精确的闭合椭圆，并且轨道的近日点也是固定的（近日点是行星轨道上离太阳最近的一点）。但是水星轨道的问题是，它的近日点不是固定的。其他行星的引力，以及太阳系里小行星带的引力，加在一起使水星轨道受到一个很小的附加影响，它使得轨道产生进动，亦即近日点随着时间逐渐“前移”，在300万年内移动一周（图6）。但是，除了所有已知的引力影响外，还有一个完全解释不了的附加进动——所以称为“异常进动”——根据天文学家们的观测，它仅仅是每世纪43弧秒。在爱因斯坦以前，这个异常进动被认为是由一颗未被发现的行星引起的。但是爱因斯坦用广义相对论产生的时空曲率，算出了这个附加的进动值，正好是每世纪43弧秒。近来，其他一些行星的这种近日点“异常”进动也被测量出来。在观测误差范围之内，它们的值也同样

与广义相对论算出的值相吻合。

爱因斯坦马上算出来的第二个结果，是他在完成广义相对论之前就曾预期的一个效应。这就是光的轨迹被物质所弯曲。从狭义相对论以及它的基本原理之一——即光速对所有观测者都相同，不论他们的速度如何——可以得出一个推论，这就是能量和质量等效。这样一来，一束光的能量就对应着一定的质量，也就可以受到其他物质的引力作用。因此，在一个大质量天体的附近，例如在一颗恒星的附近，光线就会发生弯曲（图7）。以前，爱因斯坦也计算过遥远的星光在太阳附近发生的偏折角度，但当时他根据的是某种狭义相对论和广义相对论的混合方法，其中时空仍然假设是平直的。现在他把这重新计算了一遍，但是应用了时空的曲率。他发现新的结果正好是原来结果的两倍。现在光线必须沿着弯曲时空中的测地线传播了。

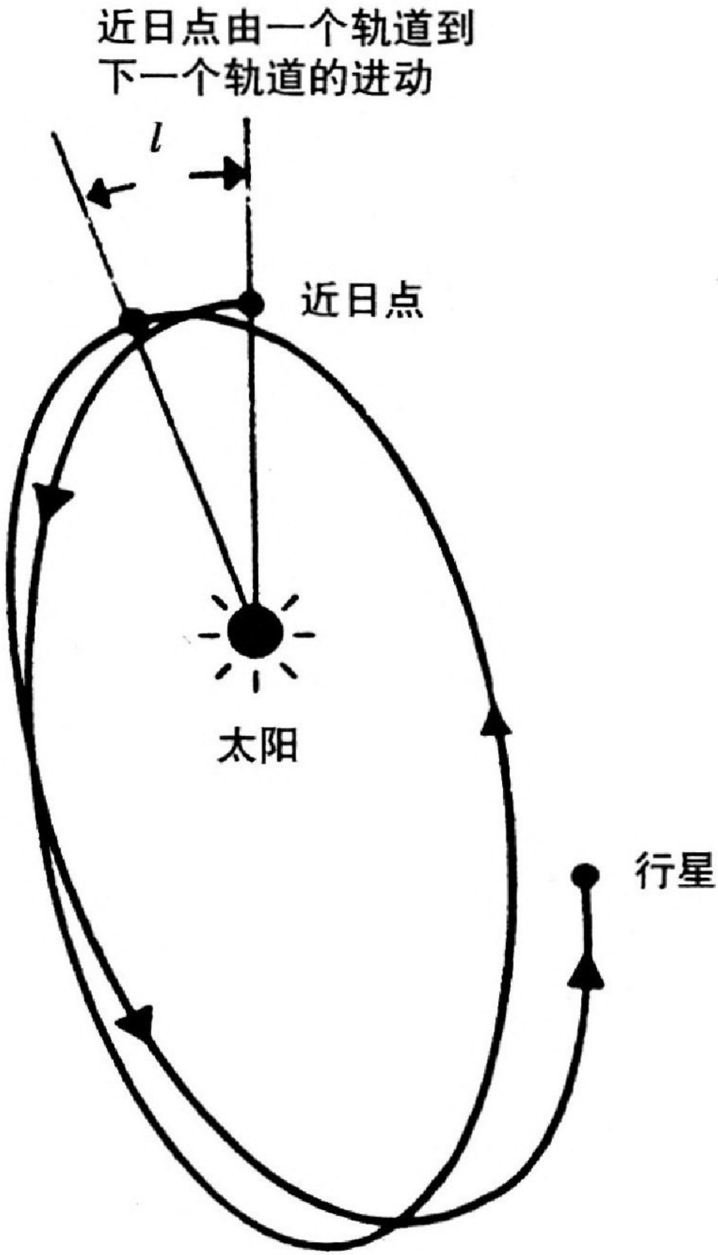



图6　行星绕日运行时近日点的进动。[录自W·仑德勒（W.Rindeier）《相对论精义》第145页。]

英国的爱丁顿帮助验证了爱因斯坦理论的第二个预言。当爱丁顿从中立国荷兰的德西特那里第一次听到爱因斯坦在柏林的工作后，他不顾当时英国和德国已经处于交战状态，就为验证这一理论做出了自己的贡献。他是教友派的信徒，这个教派从道义上反对战争，因而他被准许免服兵役，条件是继续从事他的科学研究，特别是准备监测一次即将到来的日食。1919年的这次日食，能够观测到星光从太阳近旁经过，因而可以测定光线是否发生了弯曲。通常情况下，太阳光的强烈照射使我们看不到星光。然而，从几内亚湾的普林西比岛回来以后——在那里可以对日食作最好的记录，除非是遇到坏天气——爱丁顿在皇家天文学会的一次聚餐会上，模仿奥玛·哈央姆的诗体，即席朗诵道：

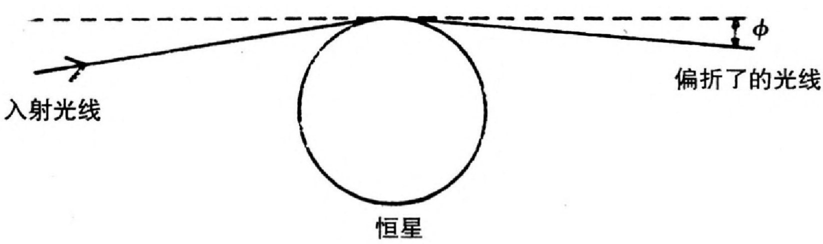



图7　入射光线掠过一个恒星边缘时会发生偏折，总的偏转角度2$\varphi$（对于太阳来说，$\varphi$的值是1″75，这是通过比较日全食的恒星位置和它们已知的位置而得到的，最近根据与太阳大致在一条线上的类星体的观测，也得到了同样的值）。[录自W·仑德勒（W.Rindeier）《相对论精义》第147页。]

噢，把我们的测量留给智者去评判吧，

但至少有一件事已经搞清——光是有质量的；

尽管其余的事还在争论，

有一件事已毫无疑问——

光线靠近太阳的时候，并不是直线前进！

在他晚年的时候，爱丁顿把这个对于广义相对论的验证，看做是他一生中最伟大的时刻。他的这个观测，也使爱因斯坦一下子在国际上赢得了声望。

近些年来的对于广义相对论的验证，是对“双脉冲星”进行的研究。双脉冲星被认为是靠得非常近的一对老年星的核，它们都已坍缩得很小。叫它们脉冲星，是因为它们发射出很规则的射电波脉冲。这一对星互相围绕对方作极高速的转动，这样就必须用广义相对论来描述，而不能用牛顿力学。它们的“近星点”的进动，要比水星和其他行星大得多。时空曲率的扰动，也已经用爱因斯坦的方程计算出来，由此可以预言，会有引力辐射从这对星发出，因而它们的轨道就会越来越小。此外，遥远的“类星体”——宇宙中最亮的天体——发射出的电磁辐射，有时候会受到一种引力透镜的作用，这种作用是由位于类星体和我们之间的某些星系引起的：每一个星系的引力场就像一种特殊的透镜，结果在我们地球上的望远镜看来，就产生了多重像，也就是原来的一个类星体变成为好几个。

总的说来，广义相对论要求从根本上更新时间和空间的概念，这个要求不是出于人为的意图，而是出于实际需要。这种更新了的时间和空间的概念，在数学上被具体化为单一的时空结构。这一时空结构决定于物质的分布，引力本身也不再明显地存在。无论如何，这是一种处理引力问题的方法。为了使读者不至于对此感到过于枯燥，我们想在此引用相对论专家威廉斯（W.Williams）教授1924年写的一首诗，它是模仿路易斯·卡洛斯《海象和木匠》的诗体而作的，诗的题目叫做《爱因斯坦和爱丁顿》，它包括下面的诗句：

“是时候了”，爱丁顿说道，

“我们有很多事情要谈及，

像立方体、钟表和米尺，

以及为什么摆锤会摆动，

空间在多大程度上偏离铅直，

还有，时间是不是具有双翅。”

“你说时间变扭了，

甚至光线也被弯曲；

我想给我的印象是，

如果它是你的原意：

邮递员今天送来的信件，

明天它就要被寄到邮局。”

“这最短的线”，爱因斯坦答道，

“不再是那条直直的线，

它绕着自己弯来拐去，

好像一个‘8’字。

而且，如果你走得太快，

你将会到达得太迟。”

“复活节是在圣诞节期间，

非常遥远就是近了，

二加二也大于四，

还有，过了那里就是这里。”

“你也许是对的”，爱丁顿说，

“但是它看来的确有些稀奇。”

## 引力时间膨胀

时间在时空中是如何流逝的呢？到1911年的时候，爱因斯坦就已经认识到，引力场越强，钟也就会走得越慢：钟离一个大质量天体例如太阳（或是一个超密天体比如黑洞，这样的效应更强）越近，比起另一个放置得很远的钟来，它就走得越慢。这个结论是由整个广义相对论得出的，称为引力时间膨胀，它不同于我们在狭义相对论中遇到过的时间膨胀效应。

这就给出了对爱因斯坦的广义相对论的第三个检验。一个原子可以被当做一只非常简单的钟——它里面的电子以极其准确的频率绕着原子核旋转，原子钟就是利用了这一自然现象。这就为科学家们提供了一个极妙的机会，通过一次实验就可以确定，全宇宙中是不是有一个“普适时间”。然而，并不需要把原子钟送到太空中去，让它作高速运动，也不需要把它放到太阳的巨大引力场附近，去验证相对论——它们已经处在实验位置上了。按照爱因斯坦的预言，太阳上的原子（更正确地说，是离子，即带电荷的原子）中的电子，它们的振荡频率比起地球上的要稍微慢一些。振荡频率的变慢，可以在离子的辐射中显示出来，也就是辐射的波长会变长一些。这确实已经在实验中得到了验证。虽然这个效应对于太阳来说很小，但是对于白矮星来说，就变得很显著。白矮星的质量和太阳差不多，但是半径却小很多，因此它表面处的引力场比太阳要强许多倍。人们已经在地球上接收到从白矮星的离子发出的光，由于引力场的这个效应，光辐射已经明显地红化。我们把这称为引力“红移”。

与此类似，甚至于对地球上不同的地点，这个效应都可以被探测到，虽然它很微小。例如，放在美国国家标准局的一只原子钟（它位

于科罗拉多的布尔德，离海平面1650米处），比起放在英国皇家格林尼治天文台的另一只同样的原子钟（其海拔仅为25米），前者每一年比后者要快5微秒（即五个百万分之一秒）。这是因为离地球中心越近，引力场也就越强。美国马里兰大学的阿雷（Carroll Alley）所做的实验，直接地显示了引力时间膨胀。他在1975年冬天所做的一系列实验中，用了两组原子钟。在一次实验中，他把一组钟留在地面上，另一组放上飞机，并让飞机在切萨皮克湾上空9000米高度处飞行。运动引起的狭义相对论效应在实验结果中被扣除了，这就是我们前面提到过的有关双生子佯谬的效应。结果发现，飞机上的时间每小时比地面上要快几个十亿分之一秒，这和广义相对论的预测完全相符。

我们也可以想象在引力场中的双生子佯谬，它类似于狭义相对论的情况。如果双生子之一跑到非常致密的天体（例如白矮星或者中子星）上去生活，他的兄弟仍然舒适地呆在地球上，则随着时间的推移，前者的年龄增长会比后者要慢许多。注意这里也是直截了当地假定了年龄是增长的。相对论既然对时间的两个方向不加区别，因而也可以同样认为，留在地球上的那一个更快地变得年轻。跟以前谈过的一样，年龄增长的现象，是与单向、不可逆转的时间有关的，爱因斯坦的理论对此并没有给出解释。

## 宇宙学和时间

我们对于时间本质的认识，总是和我们对宇宙结构的认识密切相关的。按照哲学家玻普耳（Karl Popper）的说法，宇宙学的问题，是一个任何有思考能力的人都会感兴趣的问题。宇宙的大尺度结构，也就是星系层次以上的结构，应该用爱因斯坦的广义相对论给予适当的描述。这是因为，在这样的尺度上距离变得如此之巨大，使得牛顿的引力理论不能再适用了。虽然宇宙中的物质平均密度极其低，时空的平均曲率也非常小，但是由于距离非常大，许许多多小的局部曲率总起来就产生了非常可观的影响。爱因斯坦的广义相对论方程，第一次

让物理学家们能够系统地研究这个世界的真实面貌，并且冷静地以科学的方式思考宇宙的起源问题。

观测天文学日益积累的观测资料，使我们得以运用它们去对照验证广义相对论的宇宙学模型。然而，我们在宇宙中占据的是一个无足轻重的位置。我们的太阳是一个典型的中年星（它的年龄大约是46亿年），质量平常，位于离银河系中心30000光年处。银河系像一只巨大的旋涡状圆盘，90%的物质形成大约2000亿颗恒星，聚集在它的几条旋臂上，其他的物质则为气体或尘埃。

我们的太阳系处在一条主旋臂的一个小分叉上。由于我们所处的位置已经被限定，并且探测宇宙其余部分的手段也相对薄弱，这使得可用的有关宇宙的天文观测数据极其有限。但这种资料的缺乏却给宇宙学家提供了更为自由的天地。有些科学家打趣说，宇宙学还不如算命。确实，宇宙给我们提供了一个巨大无比的实验室，但这是一个与众不同的实验室。通常的实验方法，是要能够系统地控制实验现象，这样才可能鉴别出潜在于大自然中的规律性。但是，天文学家的实验室却不受人的控制。我们只能根据天外飞来的信息，去推测天上究竟是怎么一回事。

## 时间和宇宙的创生

虽然有这些困难，我们还是已经相当有把握地了解了一些宇宙的重要特征。在这中间，哈勃（Edwin　Hubble）1929年的发现，即宇宙不是静态的，而是随着时间在膨胀，确实使宇宙学开始成为一门受到尊重的科学。然而，宇宙究竟是注定要永远膨胀下去（开放的宇宙），还是由于所有宇宙物质之间的引力作用的影响足以使宇宙停止膨胀，然后收缩，最后导致“大坍缩”的发生（闭合的宇宙），这还需要继续研究，才可以得到一个肯定的回答。引力，意味着一个不会静止不动的宇宙——即使宇宙中的所有物质在开始的时候都处于静止状

态，引力也会毫不留情地将它们聚在一起。现在的问题是，我们仍然不清楚宇宙中到底有多少物质，而这正是确定宇宙是会无止境地膨胀下去，还是会坍缩回去的关键因素。在写这本书的时候，观测数据仍然还没有精确到足以告诉我们，宇宙未来的命运到底会是怎么样。

如果宇宙是在膨胀，是否这就意味着时间有一个开始？我们可以在不同的观测资料的基础上，估计一下宇宙的最低年龄。例如，从核合成所产生的重元素的丰度，也就是氢原子核聚变而成的重核的丰富度——如我们所知，重元素是生命存在之必需——我们可以得出结论说，宇宙的年龄至少是100亿年。就这本身而言，它并不意味着宇宙必须有一个开始。

然而，20世纪50年代中期对宇宙膨胀速率的测量表示，宇宙的年龄比地球还要年轻35亿年。这就使得在50年代之中和60年代早期，宇宙学家们中间，越来越多的人支持一种叫做“稳恒态”的宇宙模型。这个模型是由宇宙学家邦迪（He rm an n Bon di），戈德（Tho mas Gold）和霍伊尔（Fre d Hoyle）提出来的，它认为宇宙中的物质不断地在创生，从而使得宇宙膨胀时，它在空间和时间上的性质保持不变。

稳恒态理论一个直接的困难是，这样的物质创生在广义相对论中完全得不到解释。但是决定性的检验——像对所有理论模型那样——必须要与天文观测相对照。1965年，彭齐亚斯（Arno Penzias）和威尔逊（Robert Wilson）偶然地发现了弥漫全天空的微波背景辐射，他们为此获得了1978年度的诺贝尔物理学奖。他们在实验中发现有一种外来的微波噪声，于是他们非常仔细地设法消除这种噪声，包括赶走了一对在实验用的角状天线上筑巢的鸽子。尽管作了这些努力，噪声还是没有被完全消除。对这种外来噪声的性质详细加以研究之后，他们不得不得出结论说，它是一种来自银河系之外的热电磁辐射，从天空所有方向均匀地接收过来，它的有效温度是绝对温度3度。这种外来热辐射只能解释为宇宙演化极早期的遗迹，宇宙在那个时候比今天

热得多，密度也大得多。这个遗迹也就是宇宙诞生的回声。

早在1948年，阿尔弗（Ralph Alpher）和赫尔曼（Rob-ert Herman）就已经预言了这个无处不在的背景辐射的存在，他们根据的是俄国物理学家弗里德曼（Alexander Friedmann）的学生伽莫夫提出的一个模型。伽莫夫用爱因斯坦方程研究了宇宙在非常早期的状态。从这一研究中他得出结论说，宇宙在那个时候应当是非常致密的，而且极端地热。宇宙开始是一团火球，它发出的辐射随着宇宙的膨胀而冷下来，这就是彭齐亚斯和威尔逊无意中发现的宇宙背景辐射。

阿尔弗的这篇博士论文的要义，是描述出现在原初宇宙的"浑汤"中的基本粒子，如何从氢经过质子和中子的核聚变而演化成为氦。它于是成为"大爆炸"理论的经典文献。这篇论文是在1948年4月愚人节那天，发表在美国《物理学评论》杂志上的。引人注目的还不仅仅是它的主题和发表的日期：阿尔弗的博士生导师伽莫夫说服了核物理学家贝特（Hans Be-the），把他的名字也添了上去，这样三个名字的谐音正好是头三个希腊字母：阿尔法、贝他、伽马。在他的《宇宙的诞生》一书里，伽莫夫写道："从希腊字母的顺序讲，如果文章只署名阿尔弗和伽莫夫，这是不公正的，所以贝特博士的名字也在文稿付印时加了上去。贝特博士在收到文稿的复印件时并没有反对……但是后来有一个传说，说是当阿尔法、贝他、伽马理论暂时遇到麻烦时，贝特博士曾认真地考虑过把他的名字改为扎查瑞斯（Zacharias）。"

很多年以前，在一篇1917年呈交给柏林科学院、题目为"广义相对论的宇宙观"的论文，爱因斯坦也隐约地察觉到这样一个宇宙史话，其中时间有一个开端，或许还有一个终结。尽管他在发展广义相对论理论时表现出了深刻的洞察力，他还是无法接受这个史话的"创世纪"和"启示录"："大爆炸"和（可能的）"大坍缩"。他本来完全可以预言哈勃关于宇宙膨胀的发现，但是他被当时流行的宇宙观念（静止

且与时间无关）所蒙蔽了。所以，他让他的理论屈从于传统之见，引入了一个新的自然常数——宇宙学常数——这真可谓是削足适履。一个动力学的宇宙学模型的可能性，大概是爱因斯坦本来可以作而没有作的最伟大的预言了。霍金把它称为“理论物理学所错过的重大机会之一”。后来，爱因斯坦终于放弃了这个额外的常数；伽莫夫写道，爱因斯坦觉得这个常数是他一生中最大的失误。

相对论关于宇宙的诞生和死亡的全部含义，由其他人的工作得到了发挥，其中著名的是弗里德曼、德西特和比利时宇宙学家兼教士勒梅特（Georges Lematre）。弗里德曼本人第一个把广义相对论作为一个自成体系的理论接受下来，并且把它用于宇宙而且得到了一些结果。事实上，利用广义相对论，弗里德曼在哈勃的研究好几年以前，就已经预言了一个膨胀的宇宙。弗里德曼的宇宙学模型意味着，如果在过去某个时刻宇宙中的物质密度是无穷大，则那时候的时空曲率也应当是无穷大。宇宙必定是从这种无法描述的致密状态中，以某种大爆炸的形式显现出来；在此之前简直是什么都没有——没有时间，没有空间，没有物质。

这样，在大爆炸之前物理定律便失去了意义，时间本身也停滞了。按照霍金的看法，在广义相对论中，“时间仅仅是一种标志宇宙事件的坐标。在时空流形之外，它便不再具有任何意义。”“问到宇宙在开始之前是什么样子，就像问到地球上北纬91°的一个点一样；它恰恰没有定义。与其讲宇宙的创生以及可能走向的末日，倒不如说：‘宇宙就是这样存在。’”

大爆炸和大坍缩在很多方面令人讨厌。它们是数学家们熟知的“奇点”的形象比喻——奇点就是体积为零、质量无穷大的时空点。相对论理论所依据的整个数学体系，在无穷大的物质密度条件下变得完全失去意义，这表明，这一理论在奇点面前不再有效。说广义相对论是时空引力的终极理论，是有一个严重的弱点的——这就是，这个理论的适用性到奇点为止。正如华盛顿大学的威尔（Clifford Will）

所指出的："认为时空奇点会存在，在那里广义相对论和其他所有物理规律都失效，是非常令人困扰的。如果物理学家不能从某些给定的初始数据而预言未来，他们就会觉得非常不舒服。而奇点恰恰是这种情况，因为从奇点出现的物理学是不受任何约束的。"

相对论过去是、现在还是充满生机。霍金和彭罗斯在1965～1970年所做的开创性工作表明，如果宇宙的行为由广义相对论的方程所决定，在过去的某一时刻，就必定有一个上面描述过的大爆炸奇点。因为物理学在奇点完全不可能给出任何描述——这是由于数学在奇点整个瓦解——这说明，我们实在不能够希望用广义相对论去处理空间和时间的诞生。然而，按照牛津大学的数学家、第一个从事奇点研究的彭罗斯的说法，这远不是意味着我们必须把广义相对论整个地抛弃。"某些人说，奇点向你表示，广义相对论是错误的。但是广义相对论的力量正是在于，它可以告诉你它本身的局限性。"彭罗斯和其他人一道，致力于把广义相对论的这一短处转化为长处，我们在第五章会看到他们是如何做的。

## 黑洞，宇宙监察和时间弯曲

引力的吸引使所有的物质受到拉力，这就使得时间会有一个终端，正像大爆炸的奇点被认为是时间的开始一样。对于质量足够大的恒星，引力可以超过其他使物质相互分离的力，而最终不留情地导致坍缩。引力场然后可能会变得如此之强，使得光都不能够逃逸，并且时间膨胀也会达到这样一个极端，使得时间看上去像停滞了一样。这样的超密天体的极限就称为"黑洞"，它是根据所谓的"事件视界"而定义的。事件视界不是一个物理的表面，而是代表任何被拉进去的物体都不能够再出来的地方。

美国理论物理学家惠勒（John　Wheeler）1967年在纽约的一次学术会议期间创造了"黑洞"这个词来描述这样的单向行为。但是黑洞

的概念早在18世纪，在一个名叫米歇尔（John Mitchell）的天文爱好者的作品中就可以找到。他根据当时流行的光的微粒说进行推理，认为光应该被引力所吸引。现在许多天文学家认为，黑洞存在于类星体和其他大的星系的核心部分。在某些有X射线辐射的双星系统中，据说也探测到了恒星质量的黑洞的存在，这是目前有关黑洞的最好的观测证据，虽然还没有一个黑洞被确凿地证认出来。毕竟黑洞是不能直接看到的，只能通过它对于其他物体的引力作用而间接地探测到。一个黑洞的事件视界从外表上看来并没有任何显著之处。一个倒霉的宇航员，也许会随着其他什么东西一起被吸进黑洞，然而他却看不到有任何特殊的事情发生；特别是，他自己的表仍然像往常一样地“滴答”走时。但是，一旦进入到事件视界以内，任何东西都不能够再逃逸出去（如果我们忽略量子力学效应的话）。并且这无法停止的引力会继续它的作用，把这个毫无觉察的宇航员拉向他自己的“局部大坍缩”点，也就是爱因斯坦方程的另一个讨厌的奇点，他头部和腿部的引力差异会把他整个人撕裂。

假定我们的太空人达姆，在事件视界外边与他的孪生兄弟迪姆分别，然后，比如说在他自己的表指向一点钟的时候，进入视界。进入视界之前，当这性命攸关的时刻迫近时，达姆每隔一秒钟给迪姆发出一个信号。达姆离视界越近，迪姆接收到的两个信号之间的间隔就变得越长，当达姆到达视界时，这个间隔就变成无穷大。然后，从理论上讲，迪姆会目睹达姆在视界处永远停滞不前；达姆的表在迪姆看来决不会真正指到一点零分零秒，因为时间被引力无限地膨胀了：时间看上去已经停滞。达姆发出的光信号的强度，在迪姆看来也越来越弱而且越变越红，因为光波的波长在强大的引力的作用下被拉长了。这样，达姆就从迪姆的视野中消失而进入黑洞。值得注意的是，对外部观察者来说，位于黑洞中心的奇点，由于事件视界而被掩盖了，这个视界阻止任何光线从黑洞内部逃逸出来。

奇点就是空间和时间的尽头。广义相对论方程中还有这样的解：

太空人可以掉到黑洞里面，避开奇点而穿过一条小通道，再从一个“白洞”跑出来。“白洞”就是黑洞的时间倒转。这个特征是由于广义相对论是时间对称的理论：“白洞可能存在于另外一个宇宙，也可能存在于我们宇宙的另外一个部分。在后面这种情况下，可以利用黑洞到遥远的星系去旅行。如果星系际旅行具有现实可能性的话，我们确实需要某种像黑洞那样的东西，”霍金这样说过。然而对太空人说来不幸的是，“像宇宙飞船飞临这样的最小的扰动，也会把黑洞和白洞之间的通道挤断。”白洞所描述的情形，在时间上正好跟黑洞相反，奇点的密度无穷大的物质会通过爆炸而出现，同时发出炫目的光辐射——就像在一个局部尺度上发生的大爆炸一样。随后，奇点会裸现出来，暴露在光天化日之下。物理学家们通常觉得白洞是不现实的，会导致经不起推敲的、像霍金描述的那样的物理后果。为了处理白洞，彭罗斯引进了“宇宙监察”假说，这是一个没有理论根据的硬性规定，它一开始就禁止裸露的奇点在宇宙中出现。按照这一规定，所有的奇点都应当被事件视界所覆盖。这其实是为了排除时间对称可逆理论中令人讨厌的现象，而做出的人为假定的又一个例子。

## 人择原理

有助于我们仍然在黑暗中摸索的宇宙学，一个饶有兴趣的看法是“人择原理”。它使我们注意到这样一个事实，即我们的存在本身，在相当大的程度上决定了我们所看到的宇宙的特点。为什么宇宙是如此浩瀚，而生命却是如此罕见？我们所看到的宇宙，它的尺度确实十分巨大，大约有130亿光年。由于它的膨胀，这意味着它的年龄也应该差不多是130亿年。而另一方面，生命决定于与氢一起存在的另外一些元素，最主要的是碳、氧、氮和磷。它们不会在原初大爆炸时就已经生成，大爆炸过程中只可能生成一定数量的氢和氦原子核。重元素的生成必须要等到星系和恒星形成时期，恒星的内部像巨大的熔炉，可以使轻元素聚合在一起而引起核合成。在此以后，还需要有长达几十亿年的加热，才能够最终生成这些重元素。因此，为了人类今

天的存在，从宇宙创生以来至少要经过如此长的时间。正像巴罗（John Barrow）写道的："发现宇宙尺度如此巨大是不足为怪的，因为我们不可能生活在一个比它小很多的世界……这是一种认真的想法：即整个也许是无限的宇宙结构，和地球这样的行星上的生命演化之必需条件，居然是如此密切相关的。"

1973年卡特（Brandon Carter）第一次提出"弱人择原理"，它只不过是说，生命（我们自己）的存在也许确定了我们看到的宇宙的某些性质。它其实是沿循了惠特罗（Gerald Whitrow）1955年的开创性工作，惠特罗当时根据三维数学物理学的许多特点论证道，我们生活在一个三维空间，是和我们作为能处理信息的理性观察者有关的。惠特罗然后把一个非常大的宇宙这个需要和生命所需要的条件联系起来。这一原理引起诸多争论，因为这些年来有人提出了另一些更加富于推测性的看法。比如，"强人择原理"认为，宇宙必须是能够容许生命得以存在；而"最后人择原理"则补充强调，一旦生命存在于宇宙，它就决不会灭绝。这两种说法初看上去更像形而上学而不像科学，比起科学家使用的说法来，它们与宇宙目的论有着更多的共同点。宇宙目的论认为宇宙具有某种目的，从而被神学家所偏爱。

## 时间旅行

广义相对论还有另外一个使人极感兴趣的推论：名副其实的时间旅行。像我们在双生子佯谬中所看到的，在一个非常有限的意义上，狭义相对论和广义相对论的时间膨胀，都允许一个观测者相对另一个观测者作"时间旅行"。然而，事件的时间顺序对所有的观测者都是相同的，即使他们不能接受一个普遍适用的"现在"：在任何运动状态之下，没有一个观测者会看到，光会在从恒星发出来之前到达地球。

但是，著名的哲学家哥德尔（Kurt Godel）1949年曾表明，按照爱因斯坦的广义相对论，旅行回到过去的微妙技巧是可行的。他发现

了一个满足爱因斯坦方程的转动宇宙模型，其中返回到历史中去的旅行是可以容许的。但这里暗含着一些使人不安的结果。如哥德尔所说："这种情形包含着某种荒谬，因为它容许一个人返回到他曾呆过的一些地方的过去。他会在这些地方发现一个人，此人正是过去某个时候的他自己。这样他就可以对这个人做某件事，而此事在他的记忆中从没有发生过。"假使时间旅行是可能的，它的确会造成一些荒诞的结果。例如像哥德尔想象的最耸人听闻的故事——新生儿自谋杀。如果这样的事能够实现，那么这个人就不会活到能干这件事。这是称作"反证法"的一种逻辑论证的一个极好的例子。

哥德尔发现的模型和我们所生活的宇宙并没有任何相似之处，所以我们尽可以把它只作为爱因斯坦方程的一个没有物理意义的解，而弃之不顾。看起来，根据"反证法"，时间旅行必须在我们的宇宙中被排除掉。正像牛津大学的天文学家雷西（Cedric Lacey）所说："时间旅行在任何合理的宇宙学模型中都是不容许的（但是我猜想这正是'合理'的定义！）。"

另一方面，宇宙学家们发现时空景观中有一个奇异性质，他们把它叫做"蚯蚓洞"。蚯蚓洞的说法是惠勒首先提出来的，它也是爱因斯坦方程的一个解，它连接着一个宇宙的相隔很远的不同部分，或者甚至于把分离开的宇宙也连接起来（图8）。

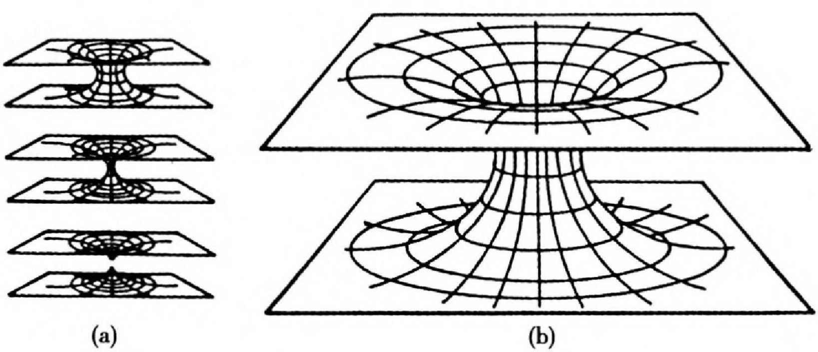



图8　两种可能的时空中的蚯蚓洞。在经典（牛顿或爱因斯坦）的情况下（a），蚯蚓洞在一

个旅行者能够穿过之前就被挤断了。在量子情况下（b），坍缩是可能避免的，这就会容许新生儿自谋杀以及其他诸多事件的发生。

一个旅行者掉进一个适当的蚯蚓洞后，就可能回到他的“过去”的某个位置。但是，牛顿力学或者爱因斯坦力学中所描述的物质，不可能支撑蚯蚓洞存在这么长的时间，以使得这个无畏的旅行者回到过去——他会在这个过程中不知不觉地被挤压得粉碎。然而，值得注意的是，最近有三个宇宙学家，莫里斯（Michael Morris）、托恩（Kip Thorne）和尤瑟福（Ulvi Yurtsever），从理论上提出了一个看法，认为如果考虑了物质的量子性质，蚯蚓洞的坍塌就可以避免，因而新生儿自谋杀及其他许多类似的事情就依然存在可能性。他们的工作是由于受到一本科学幻想小说《门路》的启迪，在这本书里，作者萨甘（Carl Sagan）描述了一种古代文明人类建造的蚯蚓洞，它能够用以实现超高速的时间旅行。然而一定会有人问道，在这推论的中间什么地方，是不是“漏掉了某种使得物理学保持一致性、并且防止我们作时间倒退旅行的基本限制”。

## 相对论的不足之处

现在回到我们的主题：爱因斯坦对解释时间箭头无能为力。这其中的原因，是和因果律的概念密切相关的，这个概念认为，结果绝不可以先于其原因。设想在一个给定的参考系中，某一事件——例如说一个击球员在空间某X处，在某个给定的时刻击中一个板球，引起晚些时候发生在空间坐标Y处的另一个事件——例如说是板球落到外场员的手里。为了使Y处的事件能够发生，一定有信息从X处以某种速度传到Y处，在我们设想的情况下，这速度就是板球的速度。如我们上一章中所看到的，这个概念在牛顿力学中具有基本的意义。它被爱因斯坦全盘接收下来，在构造他的相对论时，作为一个无须质辩的指导原则。

爱因斯坦坚持认为因果律具有不变性——对所有惯性参考系中的观测者，X都必须早于Y。这看来是合理的，因为我们知道，事实上是先击球而后接球。然而，在保持因果律的同时对洛伦兹变换进行分析就会发现，没有任何作用可以比光速跑得更快。因而光成了一种把事件联在一起的宇宙脐带。否则的话，就应该有这样的参考系（可以是板球场、瓷器店或者恒星），如果从Y发出的信号比从X发出的信号跑得快，参考系中的观测者就可以看到Y比X在时间上要早发生。用另外一种方式说，爱因斯坦狭义相对论的第二条公设，即存在一个由光速给定的极限，保证了无论观测者的环境如何，因果律都保持不变。

爱因斯坦坚信因果性的概念是物理学的基石，他这一点怎么样形容都不会过分。这其中的原因很容易懂，只要我们来设想一个违反因果律的世界。在这个世界里，一块卵石会从地上飘浮起来让你抓住；比这更糟的是，在一块石头还没有落下来之前，你也许早就被它击倒了；或者，你会在还没有出生之前就杀死了自己的祖母。因果律和决定论是密切相关的，我们可以回忆一下，牛顿力学是决定性的——给定了某一时刻的动力学条件，就足以预言该体系以后的所有行为。爱因斯坦对动力学的修改并没有触动决定论。而且，因果律在相对论中起着更加主导的作用，因为它首次考虑了信息传递的有限速度。在牛顿物理学中是不言而喻地假定，像大质量天体之间的引力吸引这样的作用——也就是大距离上的作用——是瞬时之间传播过去的。

爱因斯坦对常识上的因果观念的坚持，遗憾之处在于它是自我毁灭的。在第二章中我们已经看到，牛顿对于世界的描述是一种严格的决定性的描述，只要提供足够的信息，将来和过去都可以构造出来。我们也看到他的描述是可逆的——在向前的时间和向后的时间两者之间，并没有物理上的区别。搞清楚一个系统在任何一个时刻的情况，就马上获知该系统在整个过去和将来的情况。这是由于牛顿理论的因果性结构：在某个时间点上发生的事件，唯一地导致后来某个时刻的结果；除此以外，由于运动方程的时间对称性，后来的结果也同样可

以说成是以前事件的原因，因为这两个过程都是牛顿方程的数学解，所以物理上都是可能的。

同样的问题爱因斯坦的理论中也有。他所有的相对论方程中（不论是狭义相对论还是广义相对论），大致说来都含有与牛顿力学同样的决定性结构和时间可逆的性质。向前的时间和向后的时间没有区别（这就是为什么我们必须引入某种特别的要求，例如像宇宙监察假说和禁止时间旅行）。这样，当把正的时间换成负的时间时，度规结构（它决定了时空的几何性质）以及由它而来的运动方程都保持不变。同样地，给出在任一时空坐标上的条件后，宇宙的整个历史和将来就可以计算出来了。正像牛顿力学不能解释许多我们熟悉的不可逆过程（诸如雪人的融化，杯中咖啡的冷却，雕像的风化和年龄的增长）一样，爱因斯坦的相对论也对它们无能为力。

20世纪理论物理学的另一个引人注目的成就是量子理论。在量子论中，因果关系的全部概念被从根本上重新评价了。按照量子论的习惯解释，不能再简单地认可因果性原则。量子力学最终放弃了因果律，对此爱因斯坦从来没有赞同过。但是量子力学这样做，的确可能提供某种克服上述困难的途径。

把宇宙的演化在时间上从现在起倒推回去，我们会发现接近大爆炸的时候，宇宙中的所有物质难以置信地挤压在一起，时空的弯曲也超出任何可以想象的程度。接近大爆炸奇点的时刻，我们所面临的情况是，爱因斯坦的理论必须开始失效。这是因为，即便是爱因斯坦的强有力的理论，也先天地带有一些假设，这些假设在非常短的距离上是不正确的。如我们在前面强调过的，广义相对论在非常大的宇宙尺度上是正确的，因而它替代了牛顿物理学；但是在基本粒子、原子和分子领域，经典物理学就要用量子理论来替代。奇点是十分棘手的，因此科学家们力图寻找一种途径，把量子论和相对论综合起来以绕过奇点。

这样的一种综合，是否就是时间箭头产生的关键？彭罗斯认为是的。我们将在第五章中看到，他相信“量子引力”是一种内禀的时间不对称的理论。然而他又说，“它仍然是我所担忧的事情之一。答案看起来几乎是显而易见的，但是当我这么讲的时候，看来没有人同意我的说法。”也许，一个更能令人满意的时间概念，可以通过把相对论与量子力学合二为一而得到。量子力学支配着魔幻般的微观世界，现在我们就来谈谈它。

# 第四章　时间的量子跃迁

如果原子研究不能适合空间和时间，那它整个目的就失败了，我们也就不知道它的作用究竟是什么。

——薛定谔

《致维里·维恩》，1926年8月25日

对于时间箭头的存在，量子力学给了我们饶有兴趣的启示。如我们将会看到的，它认为时间的流逝是由某种非常简单的事情决定的：即我们自己对于变化的观测。它揭开了原子世界的奥秘，显示了有一种极其微小的粒子（长寿K介子），其存在表示时间是不可逆的。但是物理学家们仍然在为此争论不休：这确实是一个基本线索呢，还是一种风马牛不相及的东西。有一件事倒是很清楚——量子世界中到处都是问题和佯谬。例如，这个新理论在许多方面仍然在步它的前任的后尘，时间似乎是既可以向前又可以向后。它认为事件会无休止地重复出现，但同时也支持这样的观点，即锅里的水决不会自发地沸腾。它认为，一只猫在同一时刻既是活着又是死了，而且有些东西在同一瞬间，既是无处不在又是无处在。它是一种如此奇怪的理论，许多帮助它创立的科学家——其中包括爱因斯坦——后来极力要与它脱离关系。在它创立了几十年之后的今天，对于量子论究竟意味着什么，仍然有许多不同的看法。

量子论涉及的是物质在最微小的尺度上的性质，这其中包括原子，它是化学元素的最小单位。在试图描述世界在这种微观层次上的行为时，我们发现牛顿力学不能用了。和它在相对论涉及的高速大质

量物体情况下的失效相比，牛顿力学在处理微观世界情况下的失效更为明显。与此相反，量子论却在原子层次取得了非凡的成功。我们对化学反应、激光、晶体管和作为现代计算机技术基础的二极管的详尽知识，都依赖于量子论。今天，原子的存在看来是没有争议的了——原子和分子的图像甚至可以借助于场离子、电子或者扫描隧道显微镜而看到（见彩图页）。但是人们很容易忘记，对原子存在的争议其实还是不久以前的事。虽然原子论的思想古代就产生了，但多少个世纪以来，它一直受到压制。

原子的史话也许始于大约公元前500年爱琴海的一个海港阿布德拉。两位原子论的先驱者，一位是哲学家卢西普斯（Leucippus），另一位是他的学生，阿布德拉的德谟克利特（Democritus of Abdera）。他们的观点与现代科学观点并没有太显著的差异。他们认为，世界是由微小的、看不见的而且不能够再缩小的物体所组成——这些物体只是在外形和大小上有区别——它们在无限的真空中处于永恒的运动状态。他们把这种物质实体称之为原子，意思是不可再分的，并且认为一切物体，从桌子到海龟，都是由于原子的偶然碰撞而形成的。原子论者还用原子来解释感觉现象，例如味觉和嗅觉。不幸的是，由于柏拉图和亚里士多德的影响，原子论被人们遗忘了。这几位西方哲学之父主张，物质可以被无限地分割，不存在不能再被进一步分割的最小单元。原子论于是被打垮，在阴影之下度过了2500年。

为什么原子论又东山再起了呢？这主要归功于一位名叫道尔顿（John Dalton）的教友派教师，他1766年出生于昆布兰郡的依格列斯菲尔德城。他在1808～1827年写的题目为“化学哲学的新体系”的两卷体专著，使原子论得以新生，并且成了现代化学的奠基著作。道尔顿认识到，原子有助于解释越来越多的科学现象，包括气体的行为和一种物质到另一种物质的化学变化。道尔顿认为，原子是物质最小的不可再分的单元，并仍然具有这种物质的化学性质。他主张，化学反应只不过是这些物质的基本“砖块”的分离和组合。今天我们通常把

这些“砖块”称为分子——它们是原子可以参与化学反应的最低组合。例如，水分子就是由两个氢原子和一个氧原子组成的。

开始的时候，其他化学家对道尔顿的主张将信将疑。他们了解他的想法，但是并不认为原子确实存在，所以只是把原子作为一种方便的工具，用来解释他们的实验数据。法国化学家杜马斯（Jean Baptiste Dumas）甚至说，“假使我能做主的话，我会把原子这个词从科学上抹掉。”然而过了一段时间，化学家和物理学家们开始认识到，他们已经积累了许多独立的证据，这些证据毫不含糊地倾向于原子论。当时，争论的焦点主要是气体和所谓的动力论，即用原子和分子来解释气体性质的理论。物理学家们，像麦克斯韦和玻耳兹曼，提出了简单的模型来解释气体对容器的压力。他们把气体形容为像台球那样的一群刚性球的集合，它们不停地快速撞击容器的器壁，这种碰撞过程可以用牛顿力学来描述。气体的性质用构成气体的原子和分子的运动来解释。压力可以很容易地从刚性球碰撞容器壁的速率计算出来。热是分子快速随机运动的结果：气体越热，分子的运动也越快。

但是对于像马赫和德国物理化学家奥斯特瓦尔德（Wil-helm Ostwald）这样的死硬派原子论反对者来说，这些还仍然不足以说服他们。作为实证主义者，他们强调说，谈论一个无法直接看到的世界是毫无意义的。原子论者所需要的，是能够直接展现在怀疑者眼前的分子作用事例。到1905年他们认识到，有一个事例早就可以用了，它在道尔顿那个时代就已经被发现。这就是所谓的“布朗运动”——悬浮在水中的很小的花粉（以及尘埃或煤烟）颗粒，像跳舞那样的运动。早在1827年，苏格兰植物学家布朗（Robert Brown）就曾经在显微镜下观察过这种作用，但是对此一直没有令人信服的解释。直到爱因斯坦，才对这个问题的研究做出了独特而卓越的突破。他解释说，布朗运动是由于悬浮的颗粒，与它们周围看不见的水分子的随机碰撞。

这是物质原子论的一个有力证明。但是那时候经典的原子概念——像卢西普斯和德谟克利特所设想的那样——已经过时了。它已经

在19世纪将近结束的时候，被放射性的发现所取代。1895年，德国物理学家伦琴（Wilhelm Rontgen）偶然间发现了一种神秘的射线，他把它叫做X射线。第二年，法国的贝克勒尔（Henri Becquerel）在研究X射线的时候，探测到有很强的辐射从铀的化合物中发射出来。由波兰化学家玛丽·居里以及其他人所做的后继工作，把这些零散的发现汇总到一起，从而发现有些元素的原子，可以衰变为化学性质完全不同的其他元素。放射性元素的这种变化——几乎类似于中世纪的炼金术士们所梦寐以求的——在1902年由卢瑟福（Ernest Rutherford）和索迪（Frederick Soddy）用定律的形式清楚地表述出来。从这一点上说，现代的物质原子论已经同古代沿袭下来的观念断绝了关系，因为现代原子论表明，原子本身具有结构，而且可以被进一步分割。

## 原子结构与经典物理学的失效

在20世纪的头10年间，接二连三的有关原子结构的发现，就像是一场竞赛。无疑这其中最重要的实验，是盖革（Ha ns G eige r）和马斯登（M arsden）在卢瑟福指导下，于1909年在曼彻斯特大学所做的那次实验。它的结果是非凡的，使人们第一次有了印象，一个原子看起来是什么样子；而且开辟了通向新的物质观的途径。这两个人用放射性物质产生的阿尔法粒子束（在放射性衰变过程中辐射出的带正电荷的粒子），去轰击金箔。大多数粒子直接穿过这很薄的箔片，只产生很小的偏转。然而有极少数粒子，却朝正好与粒子束相反的方向反弹回去。卢瑟福说，这是他从来没有遇到过的最难以置信的事情：“它差不多就是你用大口径的火炮去轰击一张薄纸，而炮弹却反弹回来把你打中。”他对这一散射结果的解释是，阿尔法粒子可能是被质量很大、但体积很小的原子核碰撞回来，原子核带正电荷并位于箔片上每一个金原子的中心。

卢瑟福的著名原子模型是在1910年圣诞节期间发表的。在他这个模型中，带负电的电子就像一个微型太阳系中的行星那样，围绕原子

核作轨道运行。原子的大部分是空的，原子的大小决定于最外边的电子轨道。原子核的半径（等价于太阳的半径）大约是一千万亿分之一米，电子运动区域的半径大约是一百亿分之一米。人们很快就认识到，放射性是由于从原子核内部辐射出粒子，这跟着表明，原子核也同样具有某种内部结构。

鉴于牛顿物理学的种种成就，我们自然地相信，电子的运动也应遵从牛顿的决定性的方程。如果牛顿力学对板球成立，为什么会对板球中的原子和电子不成立？用这种思维方式可以得出许许多多的预言，其中大多数是可以用实验来验证的。

然而，在应用到原子吸收和发射光线的情形时，经典方法就失效了。按照麦克斯韦的电磁理论和牛顿力学，带负电的电子在围绕原子核转动时，应当发出彩虹那样频率连续的光辐射。而实际上观察到的，是一系列不连续的、完全独立的谱线，很像超级市场上用的“条形码”，只是带有颜色罢了。更糟糕的是，如果电子以经典电动力学预言的方式发出光辐射，则电子会失去能量从而螺旋式地掉到原子核上——就像水流进排水管洞口那样。从经典理论得出的不可避免的结果是，作为物质结构基本单元的原子，是不稳定的。

## 量子论的诞生

量子力学以一种迂回的方式，出来搭救这些自我毁灭的经典原子。它始于一个并非存心要闹革命的德国物理学家普朗克（Max Planck）。普朗克的工作实际上要比卢瑟福的原子模型早10年。当时所有的理论都不能够解释一个物体的温度和它发出的电磁辐射的量之间的关系：比如，一个烧红了的火钳，当再加热时为什么会变白。当时的理论学家是根据一种理想的模型来做出预言的，这种模型叫做“黑体”，它百分之百地发出或吸收辐射。在辐射谱的红端也就是低频部分，黑体模型与实验符合得很好。对于高频部分，它预言物体将

发出无穷大的能量。这个荒谬的结果被戏称为“紫外灾难”。事实上，观测表明，辐射的密度在高频和低频端都很小，而在中间某个地方出现有一个峰，峰的位置决定于发出辐射的物体的温度。

到了19世纪90年代后期，人们把一些近似的定律拼凑在一起，用来拟合黑体辐射的实验测量结果。但是，对于辐射密度随着频率变化的规律，一直等到1900年10月19日，才由普朗克在德国物理学会的一次会议上，给出了一个令人满意的解释。他的这一个历史性的宣布，其根源要追溯到1897年他和玻耳兹曼的一次争论，在那次争论中，玻耳兹曼建议他用一种统计方法去解决问题。普朗克是当时最重要的热力学家之一，他自然希望热力学能够解决黑体问题。玻耳兹曼首先创立了一种统计力学方法作为热力学的基础，这种方法的根据是假设原子和分子存在。作为一个原子论的反对者，普朗克当时拒绝用这种方法。其实早在1891年的一次偶然见面中，玻耳兹曼就曾对普朗克和奥斯特瓦尔德谈到，在他看来，“没有理由认为，能量不是分成一个一个‘原子’的”。作为一个老派物理学家的普朗克，最终反悔而接受了玻耳兹曼的建议。他最后得到的定律对黑体谱给出了十分漂亮的描述。

为了导出他的定律，普朗克确实不得不假定，电磁辐射所携带的能量是一份一份的，他把这叫做量子。他发现，像物质一样，能量也只能被分成为有限的份数，而不是无限多份。他这个工作的中心点是一个数学关系，它表明，量子的能量可以用辐射的频率，乘以一个新的基本自然常数来计算，现在这个常数就被称为普朗克常数。能量和辐射频率之间的这一简单的“普朗克关系”，实际上说明了能量和频率是同一种东西，只不过是用不同的单位来表示罢了。

## 爱因斯坦出场

普朗克认为他对于黑体辐射的解释是古怪可笑的，因为它与经典

电磁理论的教义相矛盾，所以他没有能够进一步地挖掘这一解释的更深的含义。作为一个保守的科学家，他只是把他的理论当做一种用起来方便的假设，而不是当做奥妙的真理。然而他也是一个务实的人，由于这个理论是这样卓有成效，他对它深信不疑。但是当他的理论的全部含义，后来被其他人详尽地加以阐明的时候，他还是受到了很大的震动。有人这样说，当普朗克把量子幽灵从瓶子里面放出来后，他被这个幽灵吓得要死。

再一次是爱因斯坦，他把这个理论向前推进了一大步。他对量子论做出卓越贡献是在1905年，在这同一年，他关于相对论和布朗运动的论文发表在《物理学年刊》上。实际上，正是由于他在量子论方面的这一突破而不是相对论，使他获得了1921年的诺贝尔物理学奖（这一消息是1922年发布的）。他这一成就解决的中心问题，是所谓的光电效应。实验表明，照射在固体金属表面上的光，可以使金属发射出电子。这些电子的能量不随光的强度变化，而是随光的颜色变化。这样的行为完全不能用经典的电磁理论来说明，因为按照这一理论，光的强度越大，从金属里面打出的电子的速度也就越大。但实际观测到的是，当颜色给定时增加光的强度，只会打出更多的电子，而电子的能量却保持不变。为了解释这个现象，爱因斯坦认为，能量是以微小份额的形式由光线携带的，他把这称为“光量子”。比较亮的光线表明有更多的量子——所以能从金属中打出更多的电子。频率比较高的光意味着更大的量子，所以逃逸出来的电子会具有更大的速度。在某一量子尺度下，电子就完全不能够获得足够的能量而离开金属表面。

对他这（完全基于演绎）的特殊解释，人们当时是表示怀疑的。因为尽管有普朗克早先的工作，但人们仍然普遍认为，电磁辐射的能量是连续的。爱因斯坦的建议在某种意义上是说，光是由微粒构成的，这是牛顿支持的一种观点，它早在1678年就已经被荷兰惠更斯的光的波动说所取代了。波动说看上去是如此优美，它清清楚楚地解释了一系列光学现象，例如折射、反射和干涉（当从两个光源发出的光

迭加在一起时，就会发生干涉，这时候产生明暗相间的干涉条纹），因而使人们不愿放弃。

实验物理学家们用了许多年时间，详细地检验了爱因斯坦的光电效应理论。到了1916年，它被完全证实了。这个理论的非凡成功，最终迫使科学家们在20世纪20年代重新考虑光的本质。然而，令人啼笑皆非的是，这一理论与波动理论的冲突，却使爱因斯坦在其后半生中忧虑不安。他总是强调，光量子说只是一种暂时性的假定。1951年12月12日，那时他已近垂暮之年，他写给他的朋友贝索说："这50年来，冥思苦想并没有让我接近这个问题的答案：什么是光量子？当今任何一个普通人，都认为他知道这个答案，但是他是错的。"尽管爱因斯坦本人的保留态度，光的这一独特的存在形式，今天已经被毫不含糊地证实了。光量子被命名为"光子"，这是物理化学家列维斯（Gilbert　Lewis）1926年建议的。但是这并非意味着光不是一种波动。这是因为，光子在具有粒子性质的同时，也具有波的性质。它有时表现得像波，又有时表现得像粒子。这是我们第一次遇到的量子世界的奇怪特征之一：粒子和波的双重性。

## 波和粒子

波-粒子双重性对显示光的波动本质的经典实验来说，是一个意想不到的周折。这个实验是杨（Thomas　Young）在19世纪初进行并分析的。他让从一个光源发出的光，投射到一个开有两条狭缝的不透明的屏上。这两条狭缝就像一个二次光源，光穿过它们之后继续传播，最后投射到一个屏幕上，形成明显的明暗相间的带状条纹，这是一种典型的干涉作用。这个实验是光的波动本质的一个很好说明：让一列并行的水波通过水中开有两条狭缝的屏障，也能够产生出类似的干涉图样。当水波从这对挨得很近的狭缝中通过后，有些地方的波相消，另一些地方的波相长，这就出现了动静相间的干涉条纹。

现在让我们想象，在杨做的实验中，如果我们只用一个光子，会产生什么样的结果。一个粒子的质量是集中在一个单独的点上的，而波是一种没有质量的实体，它弥散在一个有限的范围之内。这样，一个光子必然是只能穿过这两个狭缝之一。然而结果却是，如果我们把单独的光子一个接一个地向这两条狭缝发射过去，并记下它们到达屏幕的位置，最后我们会得到以前用一束光照射时那样的干涉图样。这样看起来，一个单独的光子，会由于它的波动性质而对两条狭缝都有感觉。

这种双重性质不仅仅是光才有——它含有更广泛的意义。这一点，是一个名叫德布罗意（Louis de Broglie）的法国年轻人顿然领悟到的。德布罗意出身于贵族家庭，他战时曾从事无线电方面的工作，这使得他的兴趣从中世纪教会历史转到物理学方面来。他主导法国物理学界的时间超过了一整代人（有人说这是件坏事）。他对于量子论的贡献是如此极端彻底，以致有人讽喻为“法国喜剧院”（法国资格最老的剧院，路易十四世创建于1680年——译者注）。德布罗意从光量子中认识到，正像光波可以表现为粒子一样，粒子也可以表现为波。为了使他这个直截了当的想法有一个坚实的基础，他提出了一个简单的数学关系，把一个运动物体的动量和与物体相应的波联系起来：这个“粒子”的波长反比于它的动量（即粒子质量和速度的乘积），这个比例常数再一次又是普朗克常数。这样，粒子的速度或者质量越大，它的德布罗意波长也就越短。

1923年德布罗意首次把他的想法写成三篇短文，发表在《巴黎科学院学报》上。然后他着手写他的博士论文，该论文于1924年完成。论文审稿人之一是朗之万（Paul Langevin），他显然对这一想法感到十分震惊。他把这篇论文的一份拷贝寄给了爱因斯坦，爱因斯坦马上认识到其中的意义，他在回信中说，德布罗意“已经把大面纱的一角揭开”。无疑地，爱因斯坦的这一评价对论文顺利通过口试起了影响。但是德布罗意并不希望仰仗爱因斯坦的权威——他向自己问道，

如何才能够使这一想法得到验证。1927年，德布罗意的想法终于毫不含糊地被实验证实，其中一个是戴维森（Clinton Davisson）和盖尔末（Lester Germer）在美国贝尔电话实验室做的，另一个是由瑞德（Alexander Reid）在汤姆孙（George Thomson）的指导下，在苏格兰的阿伯丁大学完成的。非常有趣的是，以前J·J·汤姆孙曾因为证明电子是粒子而获得诺贝尔奖，而此时他的儿子却因为证明电子是波而获得同样的奖。

物质的粒子和波的双重性质具有奇异的结果。让我们想象一个“量子台球”游戏。球的波动性会产生许多令人惊异的后果。无论击球人瞄得多么准——例如，他瞄准一个在球袋边上的红色球——这个球却总有机会落到球台另一端的球袋（因为作为波，这个红球应该扩展到整个桌面）。由于一种叫做“隧道效应”的量子现象，也有可能直接越过另一个球，这时被击的球可以径直地穿过另一个位于中途的球。这同样是由于球的类波性质，使它可以扩展而跨越障碍物。

如果德布罗意的想法是对的，那么为什么我们在平常的台球游戏中，看不到这样的类波作用和其他的量子现象呢？这个原因可以从德布罗意关系中找到。德布罗意关系表明，粒子的波动性决定于它们的质量——质量越大则相应的波长越小。对原子而言，这一波长相对于它们的“尺度”来说很大，而对于通常的（宏观的）台球而言，这一波长就小到了微乎其微。只有把台球游戏缩小到微观尺度，我们才能够观察到这些奇特的量子效应。

## 玻尔原子和量子论

上面说的这些，和原子有什么关系呢？为了了解这一点，我们首先必须回到1913年，去看一下丹麦理论物理学家玻尔的研究工作。对于玻尔的才智有各种评价，它们之间有着天壤之别：卢瑟福把他形容为“我从来没有遇见过的最聪明的小伙子”；而伽莫夫却认为，玻尔的

最大特点是“思维和理解的迟钝”。邱吉尔（Winston Churchill）也重复过同样的观点，他认为玻尔是一个枯燥乏味的人。然而，玻尔确实启发和主导了研究世界的新途径，这个新途径是由于量子理论而出现的。

针对经典力学不能对付卢瑟福的自我毁灭的原子，玻尔提出了一个尽管不优雅，但是带有根本性变革的解决办法，这个办法介于经典力学和现代量子力学之间。他对围绕原子核旋转的电子，简单地提出了一个人为的假定——用新的量子法则取代牛顿力学。这些法则不能真正称为运动定律，因为它们不具备任何理论基础。这些法则只是直截了当地说，电子占据着固定的轨道，在这些轨道上面它们不发出辐射。玻尔把这些轨道称作“稳定的量子态”。这个想法真可以说是无源之水、无本之术，它和爱因斯坦关于光电效应的想法有许多共同之处。

像以前认为的那样，电子在不同的轨道上运行，颇像围绕太阳的行星。但是当原子接收到电磁能量时，例如吸收了一个光子，则其中一个电子会立即跳到离原子核更远的另一个轨道上。这就解释了独特的分立谱线，因为只有当电子再跳回到它原来的轨道，即“稳定的量子态”时，才会有光辐射出来。这里我们可以看到量子体系的一个主要特征：能量不是以一种连续的方式被吸收或者辐射出来，而是只能在发生突然的量子跃迁时，按照原子的能级而改变。

玻尔把他的模型用于最简单的原子——氢。氢原子只有一个电子，围绕一个电荷量相等但电性相反的原子核作轨道运动。氢原子核的质量比电子要大得多，它称为质子，是卢瑟福在1919年发现的。玻尔运用他的量子法则，首次解释了氢原子辐射的电磁谱。现在物理学家们不仅可以搞明白，为什么氢原子的谱具有包含一系列分立频率的“条形码”结构（这些频率对应于电子的量子跃迁），而且对谱线出现的准确频率也可以做出预言（图9）。

虽然这个新理论是令人兴奋的，但它并没有能说明较为复杂的氦原子光谱。氦原子是仅次于氢的最简单的原子，它只有两个电子，围绕一个带两倍正电荷的原子核旋转。对于更复杂的原子，这一理论就显得更无能为力。除此之外，使玻尔理论从根本上失去效力的原因，是它没有解释这个人为的量子假定。这样，它只是一个不完全的，或者是一个临时性的原子结构理论。

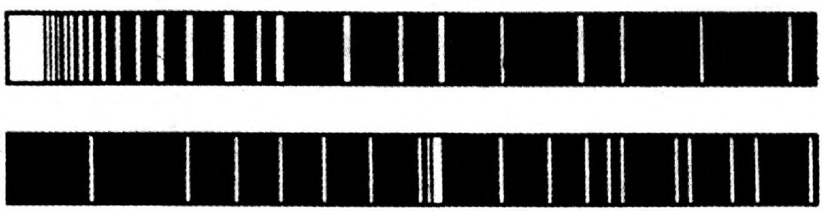


图9　原子光谱的例子。吸收线或发射线出现在分立的光线频率处，这对原子的量子本质给出了强有力的证据。[录自W.J.Moore《物理化学》第587页]

## 新的力学

下一个突破过了一段时间才来临。这个突破，是几乎同时地由两条独立的途径得到的，其中之一是年轻的德国人海森伯（Werner Heisenberg），另一个是奥地利人薛定谔（Erwin Schrodinger）。

1925年，海森伯第一个得到了一个适合微观世界的量子表述形式，他称之为“矩阵力学”。他在北海的海尔古兰岛上染上了花粉热，康复以后，他就创造出这个世界上第一个完善的量子理论。他的理论之所以叫做矩阵力学，是因为它用一种叫做矩阵的数学形式来表述微观世界。矩阵的代数和通常的数字代数很相像，但是有一个很重要的例外。在通常的乘法中，二乘三和三乘二是一样的。但是矩阵A乘矩阵B，并不等于矩阵B乘矩阵A。后来人们认识到，这个不对称的数学特点联系着这样一个事实，即仅仅是测量的先后次序不同，微观世界就可能给出不同的结果。这是量子世界所显示的许多奇特性质之一。

矩阵力学相当抽象，所以薛定谔的表述方式更容易引人注目。薛定谔受到玻耳兹曼很深的影响，他自己说过，“玻耳兹曼的学说对我起的作用，就像科学上的初恋一样，没有其他任何学说能再使我如此入迷。”薛定谔1926年对原子问题的解释，被索末菲（Arnold Sommerfeld）称赞为“20世纪所有令人惊异的发现中，最令人惊异的发现”。

德布罗意的工作给了薛定谔极深刻的印象，所以他全力投入到这个波-粒子学说的研究中。他是在阿罗萨发展了波动力学的，那是个离滑雪胜地达沃斯不远的、阿尔卑斯山中的一个地方。当时他和他的情人在一起，因为那时他和他妻子的关系，处于他们婚后的最低潮。摩尔（Walter Moor）在他的薛定谔传记中写道：“像那个黑发女郎激起了莎士比亚诗创作的灵感一样，这个阿罗萨女人永远是个谜……不论是谁可能激发起了他的灵感，薛定谔的精力变得如此充沛，这的确是富有戏剧性的。他在1年的12个月中始终保持创造活力，这在科学史上是独一无二的。”薛定谔在1925年12月的一封信中，描述了他如何在为这新的原子理论而奋斗。他说，它一旦完成之后，会是一个“非常漂亮”的理论。

1921年薛定谔的创造性达到了顶点。为了使自己不受到干扰，他常常用珍珠把耳朵塞起来。那一年，他完成了六篇有关他的波动力学的主要论文，它们都发表在德国的《物理学年刊》上。

薛定谔的理论是用微分方程来表达的，比起海森伯的矩阵方法，它是一种更直观形象的描述。从数学上看，薛定谔和海森伯的表述截然不同，但是如薛定谔本人马上表明的那样，它们的结果是完全一样的：“这好像是，美洲是哥伦布航海越过大西洋而发现的，但同样勇敢的日本人，如果航海越过太平洋也会发现这块新大陆。”由于薛定谔的理论更直观、更灵活，因而人们更喜欢用它来研究量子力学问题。

德布罗意的工作表明，玻尔有关氢原子的量子法则，即必须有一系列固定的电子轨道，可以用围绕原子核的电子波来解释。这些波称为“驻波”，它类似于共鸣时，管风琴琴管中的那种波，或者小提琴琴弦被弓拉动时产生的波。在这些情况下，谐振的出现是由于管中或弦上恰好有整数个波。与此类似，只有某些德布罗意波长可以在原子周围产生出驻波。每一个波长的值对应于一个电子的“轨道”（图10）。

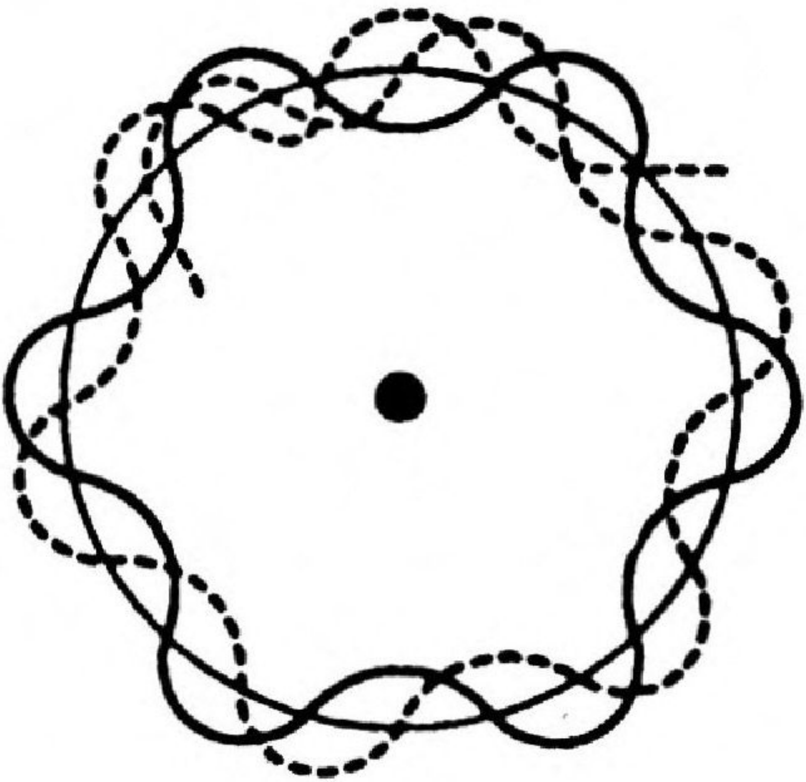


图10　德布罗意所想象的氢原子中的驻波（实践）。虚线表示非驻波将由于干涉而消失。[录自W.J.Moore《物理化学》第595页]

按照这种推理办法，薛定谔得到了一个方程，它描述物质波在微观尺度如何随时间演化。在本质上，这件工作与牛顿和爱因斯坦在他之前已经做过的一样。对牛顿和爱因斯坦来说，动力学的关键问题，

是直接地描述像台球那样的物体，其位置如何随时间而变。然而，薛定谔的工作是对于微观尺度，这时候一个粒子也同时是波。这使得分析变得更错综复杂。因此，他的方程里面含有一个全新的数学量——“波函数”，它考虑了微观粒子的波粒双重性质，并描述它们所有可能的表现。为了使波动力学和日常现象相联系，薛定谔在构造波动力学时，使得它对于像台球这样的宏观物体，就化为类似于牛顿力学的方程。于是，“薛定谔方程”就成为理论物理和理论化学中，所有方程中最基本的方程。后来的一位量子论大师狄拉克（Paul A.M.Dirac），把薛定谔的这篇论文形容为：囊括了全部化学和大部分物理学。

为了验证他的方程，薛定谔把它用于氢原子中的电子波。从应用数学的角度来看，这个方程本身并不新奇。薛定谔用标准方法解出了这个方程，他发现，跟波尔的结果一样，他的结果也和光谱实验中氢原子辐射的能量严格相符。但是现在，支配着电子、使其仅仅占据固定轨道的量子法则，清清楚楚地摆在面前。与玻尔不同的还有，薛定谔现在可以分析并成功地解释有两个电子的氦原子。

虽然有这样的成功，薛定谔却搞不清楚波函数在物理上的含义。显然，他开始时是希望能回到旧的经典概念上去，因为经典概念来源于日常生活，图像十分清楚。但是波函数使他的这一梦想破灭了。大约6个月之后，在用薛定谔方程研究像原子或电子这样的微观粒子的碰撞问题时，德国物理学家玻恩作了一个非常大胆的假设。他提出，可以把波函数解释为某种“几率振幅”，用来计算在空间某一区域发现一个粒子的几率。他认为，波函数的平方就给出了在指定地点和时间，发现粒子的几率。

玻恩的这个解释，绝不是马上就被大家接受了。实际上，对他这个解释的争论延续至今。争论的焦点是，量子力学的含义究竟是什么。答案只会有两种可能的选择，一种是薛定谔方程是一个新的基本力学的基础，而且玻恩的解释意味着，我们只可以谈及原子和亚原子

层次现象发生的几率，而不能准确地预言这些现象。这对于决定论和因果律具有深刻的意义。另一种是量子力学并不是真正基本的力学，而只不过是对我们还不完全了解的事情，做出统计说明的一种方法——这些事情在某种我们尚不清楚、更深的层次上，是严格决定性和因果性的，正如牛顿物理和爱因斯坦物理那样。

爱因斯坦正是在这一点上与新一代的量子力学分道扬镳。新一代量子力学采用上面第一种选择，而爱因斯坦却固执地坚持第二种："上帝不会把世界当做骰子"，这是他在辩论时常说的一句名言。爱因斯坦认识到，波函数的几率解释虽然得到普遍赞同，但它从根本上破坏了因果律，而因果律是许多世纪以来联系原因和结果的观念。他在为纪念牛顿逝世二百周年而写的一篇文章中说："牛顿理论的精髓可能会给我们提供力量，去恢复物理现实与牛顿教诲中最深奥的特点——严格的因果律——之间的和谐。"一件非常引人注目的事情是，德布罗意和薛定谔两个人，都对新物理学所走的方向持意味深长的保留态度。在一次与玻尔的著名交锋中，薛定谔说道："如果我们仍然不得不去建立这种该诅咒的量子跃迁，那我当初真不该和量子论打交道。"但是玻尔、玻恩、海森伯和他们的弟子们，对此异议置之不顾。他们按照自己的方式，为这一学科的"正统"说法打下基础。今天他们的说法已被广泛接受，尽管实际上，它带来的问题和它给出的答案同样多。

## 无处不在，同时又无处在

量子论所描述的世界使人感到震惊，它的一幅图像可以用图11（b）所示的一个非常简单的实验来描绘。这是一个用电子来做的杨氏狭缝实验。电子源每次向两条狭缝只发射一个电子，狭缝后面是一只荧光屏S，它用一次闪光来显示电子到达了屏幕。在这个实验中，荧光屏上会出现一幅干涉图样，上面电子数目的分布与荧光屏上的位置有关。

我们回想一下，电子既不是粒子也不是波，但是具有两者的属性。如果电子是粒子，它只会或者打中障碍屏而被撞回，或者穿过两条狭缝之一，从而引起荧光屏X处或Y处的闪光[图11（a）]。重复做这个实验，将会在X处和Y处引起同样多的闪光。另一方面，如果电子是散布在空间中的波，则波状扰动将同时经过两条狭缝[图11（b）]，正像池塘中水的涟漪相互重叠会发生干涉一样，电子波也会在屏幕上产生特有的干涉图样，呈现我们看到的强度的峰和谷。

量子力学到底预示了什么呢？像赌赛马，它仅仅给出了成败的几率，即使只有一个电子时也是如此。实际上，它给出的是电子到达屏幕上每一点的可能性。它能够绝对肯定地告诉我们的，仅仅是在什么地方不会发现电子。在这种情况下，如果把电子想象为点粒子，则量子力学就意味着，电子可以到达屏幕上任何一点，只是要除去按波解释时波的强度为零的那些地方，这些地方电子到达的几率是零。发现电子的几率随着屏幕上的位置的变化，与波的干涉所预示的波强度的变化完全一致。然而，电子最终是到达一个固定位置，而不是扩展到整个屏上。只有足够多的电子经过仪器时才能建立起干涉的图样。

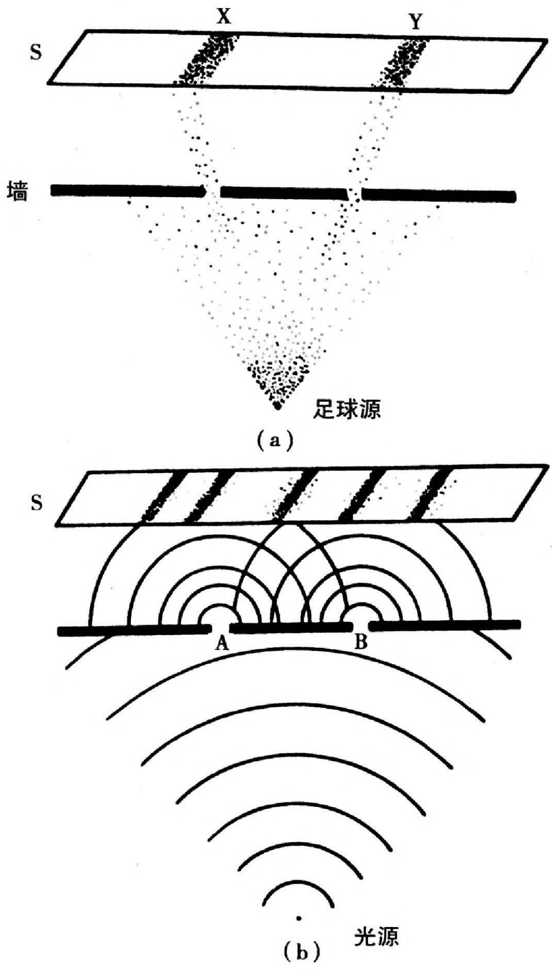



图11　（a）用经典粒子——足球——做的双缝实验，足球由源点踢出，穿过开有两条狭缝的墙而落到屏S上。球只能够达到X和Y处。（b）由一个波源发出的波，穿过开有两条狭缝的屏障，在屏幕S上造成的强度分布图样。屏幕上深色的区域相应于高的波强度，浅色的区域相应于低的波强度。缝A和缝B起着二次光源的作用。[录自J.D.巴罗《世界里面的世界》第134页。]

能够更清楚地显示这一点的量子力学的一个优美的表述形式，是美国理论物理学家费曼（Richard　Feynman）在20世纪40年代提出的，当时他是普林斯顿大学的一名研究生。通常把这种方法称为“经

历求和法”，它表明对物质世界的量子描述，当从微观领域进入到宏观领域时，是如何逐渐演变为牛顿的描述的。在费曼的描述中，一个量子粒子，比如一个电子或者一个光子，是尝试着通过源点和图11所示的屏幕到达位置之间，每一条可能的经典路径或轨道。因为粒子具有一个相应的德布罗意波长，每一个经典的“经历”都与其他的“经历”发生干涉，正如我们看到过的水波之间的干涉一样。这就导致了图11所示的微观粒子干涉图样的几率分布特征。但是如果使粒子的质量增加，比如增加到像足球那么大，则这种方法表明，相互抵消的干涉几乎在粒子所有的路径或“经历”上发生，只是严格地在牛顿力学所预言的路径（轨道）上面，不会发生这样的相消干涉。

如我们已经说过的，量子论不能够预言电子将到达哪里：它只是给出事件在给定地点发生的几率。这样，如果只有一个电子向狭缝发射过去的话，量子力学所能告诉我们的，是这个电子将会在荧光屏某个地方引起闪光，但波强度为零的那些地方除外。这种可能性如何变成现实性呢？闪光可能在屏上多处地方出现。然而，几率转变为确定性是由测量实现的：一旦在某个特定的地点发现电子，在其他地方发现它的几率就完全降到零。只有在实验重复大量次数的情况下，几率分布才变得有意义，也才能得到干涉的图样。

如果认为量子力学给出了最基本的描述，那么询问电子的行踪就没有意义，除非电子已经打到了屏幕上。因此我们只好得出结论说，电子是以某种方式扩散在空间和时间之中，它从两条狭缝中都穿过并且自己与自己发生干涉，直到最后奇迹般地瞬间瓦解在屏幕上某一点处，这地点完全是随机的。因而，我们可以说，电子是无处不在，同时又是无处在。

## 哥本哈根解释

玻尔大胆地正面处理了测量带来的难题。他的方法被大多数物理

学家所采纳，被称为量子力学的“哥本哈根解释”，因为他一直在这个城市生活和工作。这一解释的基本前提是，我们对微观世界的描述受到我们语言贫乏的限制，而语言是建立在经由感觉传递过来的信息的基础上的。世界具有一个经典的部分，它由测量行为所构成；同时又有一个量子部分，这就是我们正在测量的东西。换句话说，我们所观察的世界看来是独立的实在，但它仍然是悬浮在某种“非实在的”微观世界之上。这种限制是避免不了的，因此我们不能希望去给出量子过程的一个真实的描述。玻尔的亲密合作者海森伯认为：“希望有新的实验能使我们返回到时间和空间上客观的事件，大概就像希望在没有探测过的南极区域找到世界的尽头一样，完全是梦想。”实际上，我们用以了解世界的观测必然是宏观的，我们也只能用这种宏观观测来讨论。按照玻尔本人提倡的极端说法就是，这种量子力学的解释认为，凡是不能测量的现象就没有客观的存在：“不存在量子世界。只存在抽象的量子描述。”

仅仅建议“基本粒子”、原子和分子没有独立的存在，就遭到了大多数物理学家、化学家和分子生物学家的诅咒。在这些确信原子实际存在的化学家和分子生物学家看来，这个建议无异于纯粹的左道邪说。然而，他们的模型所最终依据的量子理论，却不支持这样的原子实在的观念。伯克莱主教有一句名言，他否认牛津大学新学院的一棵树的存在，因为他是背对着这棵树。量子理论其实是伯克莱这种观点的现代版本。

按照哥本哈根解释，每一个观测必须由它的宏观环境所决定。测量过程本身在量子论中明显地具有相当大的重要性，而在经典物理中，它是被完全忽略的。

## 海森伯不确定性原理

在海森伯对量子力学的发展中，出现了另一个使人惊奇的特点。

这就是他1927年提出的著名的“不确定性原理”。这一原理断言，自然界中存在一个测量精度的极限。设想一个像电子那样的物体从空间中飞过。按照经典物理学，它具有位置和动量，这两者可以被同时测量。而海森伯的原理简单说来，就是表明在亚原子领域，不可能同时精确地知道电子的位置和动量。如果想测量出某一时刻的准确位置，则它的动量（或者等效地，它的速度）就不能确定，反之亦然。

这个原理反映出波粒二象性佯谬：位置完全是一种粒子的典型性质，而波却没有准确的定位。波的特性知道得越多，可以谈论的粒子属性就越少。为了理解这一点，可以想象，当测量一个电子的位置时，会发生什么情况。例如，我们可以利用一个光子从电子处反射回来而做这个测量。虽然我们可以在某种确定的程度上，从光子最终的轨迹来推断电子的位置，但是在这个过程中，我们已经把数量不明的动量从光子转移到了电子身上。日常生活中也有类似的例子：测量一只轮胎的压力时，必然会使得一些空气逸出，从而使压力发生了改变。

总的说来，不确定性原理意味着，我们对一个量测量得越准，则另一个“共轭”量的不确定性就越大。把这两个不确定性联系起来的常数，又是我们的老朋友普朗克常数。因为它是如此之小，所以对于宏观物体，例如像台球或者牛顿的苹果，它实际上相当于零。因此，对于牛顿（以及相对论）力学所描述的物体，同时测量位置和速度并不受到限制。

海森伯的原理对时间的测量产生了一个后果。正如在任何情况下，我们不可能同时知道一个亚原子粒子的位置和动量一样，当我们在一段给定的时间间隔内测量能量时，这也有一个对测量精度的限制。按照不确定性原理，能量与时间之间的关系，和位置与动量之间的关系是一样的。对一个处在某一特定量子态的原子，能量的精确测量，必然要以原子处在这个量子态的时间——也就是它的寿命——的不确定性作为代价。反之，如果它的寿命已经知道得很准确，则它的

能量就很不确定了。能量-时间的不确定性原理，对于宇宙学会有重要影响，我们稍后会看到这一点。有些人认为，这可能就是时间如何能够开始“滴答”的关键。

## 玻尔-爱因斯坦论战

爱因斯坦不同意不确定性原理是自然界的一个基本事实。如我们已经谈到过的，他宁可相信量子力学是一种数学手段，用以在统计意义上预言大量实验的结果，而不是对单一实验可能结果的最好描述。这导致了爱因斯坦和玻尔之间关于量子力学基本原则的一场著名论战，他们各自对于这一原则的看法，注定了他们此后的一生。有一次，在1930年在布鲁塞尔举行的第六次索尔维会议上，爱因斯坦提出了一个不用在实验室中进行的“思维实验”。他设计这个实验，是为了反驳能量和时间之间的测不准关系。但是，经过彻夜未眠的思考之后，玻尔设法借助于爱因斯坦最重要的发现之一——相对论，击败了爱因斯坦的挑战。但是故事远远没有结束。在玻尔1962年去世的第二天早晨，在他家的黑板上，人们发现画有一幅爱因斯坦1930年“思维实验”的图。看起来，玻尔直到生命的结束，还在同爱因斯坦的主张奋斗。

在开始的时候，爱因斯坦只是简单地认为新量子论是不正确的（自相矛盾的）。但是他在与玻尔的论战中屡屡败北，使他改变了他攻击的重点：他转而认为量子论不是不正确的，而是不完备的。他反对的理由，更多的是基于量子论明显地缺乏因果性，以及它与相对论原理的不相容。虽然他们之间的相互尊重并没有因为论战而动摇，但是爱因斯坦从来没有能够说服玻尔。这场较量折磨着他们两个人。有一位同事，也就是前面提到过的派斯，这样形容过玻尔有次在普林斯顿表现出来的苦恼：“他（玻尔）处在愤怒的绝望之中，不停地说‘我对我自己烦透了’……他们总是陷在关于量子力学意义的争论，并且一直到最后，玻尔也没有能够使爱因斯坦信服他的观点。”

这场大论战——爱因斯坦在其中扮演的是反对改革的角色——充满了讽刺性。我们已经看到，爱因斯坦在20世纪初期是如何因为他对自己的物理观念的自信而被孤立，尽管这种观念与当时成立已久的牛顿式传统观念完全冲突。由于他对光电效应的解释，他曾是量子理论的带头人。但是当新的力学在20世纪20年代中期破土而出时，他便不在带头人之列了。新理论的整个构思，完全和他的见解相对立。实际上一直到最后，由于他的自信，这个在把理论引导到现实方面做过诸多贡献的人物，却始终远离不断有所发现的现代物理的大进军。正像派斯所指出的，爱因斯坦对于量子论的观点“使他的形象从远远走在时代前面，一变而为处在时代潮流之外”。

然而，今天爱因斯坦也许有了雪耻的机会，因为有越来越多的科学家，对量子力学的基本原则提出了置疑。可能是由于时过境迁，离量子力学的奠基人和鼓吹者那个时代已过去超过一代人的时间，所以当今许多一流人才，并不是简单地接受传到他们手的那些卓越非凡的传统教条。在这些置疑中，一部分是围绕着探索对物理过程中时间作用的令人满意的解释。

## 时间：失而复得

初看上去，量子力学像它以前的理论一样，大大地削弱了时间箭头的基础。与像动量和能量这样的量比较，时间在量子力学中处于二级量的地位。利用称为“算子”的特殊数学工具，可以从波函数中提取可观测的信息。例如，从描写一个电子的波函数中，用这种算子可以得到位置或动量的值。但是并没有关于时间的算子，因为在量子力学中，时间并不被看做是一个“可观测量”（也就是可以测量的量）。由于这个原因，能量-时间不确定性关系的地位就有些含糊。但是，如我们将在第八章中看到的，如果时间之箭可以在一种广义量子力学中产生出来，则定义时间算子就变得可能了。

薛定谔的波动方程是决定性的，正像牛顿和爱因斯坦的运动方程一样。给定波函数在某个时刻的值，就可以严格地推断出任何或早或晚时刻的值。这个方程所描述的行为在时间上是完全可逆的。想象有一个特别的波函数，它在数学上代表一个没有观测到的电子的行为。这个波函数储存了有关这个电子命运的所有信息，一旦我们用某种测量手段去进行观测，例如用一个荧光屏，电子的行为就立刻显现出来。与此相同，这个方程使我们能够预言，如果我们在将来某个时刻进行观测，电子所有可能的行为。更重要的足，它使我们能够推断，假使我们在过去的某个时刻进行观测，电子当时所有可能的表现。只要我们仅仅是谈及几率和可能性——这一理论的看家本领——量子力学就是纯粹时间对称的。

在量子力学中同样也有一种斯多葛学派的永恒循环。量子理论运用于一个孤立系统时，会出现一种很强形式的（第二章中描述过的）庞加莱回归，这看来支持循环时间的观念。给定足够长的时间，一个孤立系统的波函数，例如宇宙的波函数，就可以回到它的初始状态。这样，量子力学并没有提供一个令人满意的基本原则，来解释时间的客观流逝，或者至少能把时间有意义地区分成过去、现在和将来。

当然，我们必须超越波函数——因为它只是包含一个系统的所有的潜在行为，而去发现在实验中实际发生了什么，也就是说，我们必须进行测量。从这一点上说，量子力学需要时间箭头。让我们回到外观上时间对称的波函数上来。当做出一个特定的测量时，会记录到电子已经到达某个地点，而且仅仅是到达这一个地点。这样，波函数——以及系统本身——必定在进行测量时经受了某种瞬时的转换。从一种反映所有可能结果的形式，变成只相应于实验中记录到的单一值，这是一个不连续的收缩。

这种从无数潜在的结果到观测结果的转换，称为波函数的“约化”或“坍缩”。如果我们采用量子力学的哥本哈根解释，可能发生的结果的数目会有无穷多个，但是当我们突然“碰”波函数时，其中只有一

种结果变成为现实。想象你坐在一座量子剧院。这有无穷无尽、丰富多彩的剧目可能会上演，从莎士比亚到考沃德（Coward），到易卜生再到威尔德（Wilde）。但是一旦大幕拉开，剧院波函数坍缩，出现在舞台上的却是克里斯蒂（Agatha Christie）的“捕鼠夹”。

这样看来，如果我们背过脸去而且不偷看，波函数将以可逆的并且决定性的方式演化。然而对电子在屏幕上位置的一次测量，就会把波函数的行为改变成不可逆的。当波函数发生坍缩时，所有这许多的可能性就收缩为单一的现实结果。在系统过去状态（潜在性）和现在状态（现实性）之间的对称性，因而就被取消了。的确，如果试图从一个给定的测量结果去反推过去，就会得出不正确的结论。这样，测量操作本身，就把时间箭头引入到量子力学描述的现象中去了。

然而，无论是坍缩了的波函数，还是原来的波函数，都没有给出时间的方向。大多数物理学家，只是简单地采用数学家纽曼（John von Neumann，出生于匈牙利）所提出的一个补充假定，即波函数一观测就坍缩。坍缩的机制并没有给出来。确实，薛定谔方程本身显然并不能描述这样的坍缩，因为方程是可逆的而且是决定性的，而坍缩是不可逆而且是随机的。这就是测量问题的要害，它对于时间箭头具有极大的意义，并且引起了许多佯谬。

## 量子猫佯谬

这些佯谬中最著名的是薛定谔的“猫佯谬”。许多物理学家以恼怒的心情看待这个佯谬，因为他们认为这个佯谬并不具有任何“真实的”后果。例如，霍金有一次说过：“当我听说薛定谔的猫时，我就跑去拿枪。”所讨论的问题如果太深奥，这种不屑一顾的态度往往很普遍：很多人在一层薄冰上仍然轻松愉快地滑行，完全漠视脚下潜伏的危险。这样，猫佯谬就被认为是从量子餐桌上掉下来的碎屑，解决它是哲学家的事，而哲学家们也确实为此大动起脑筋。然而，也许有可

能设计出实验来验证这个佯谬。

薛定谔所设想的“思维实验”如图12所示。一猫蜷伏在一箱子里面，而箱子中放有某种放射性物质，以及一个盛有致命的氰化物的小玻璃瓶。放射性衰变本身是一种量子过程，因此它的发生只能在几率的意义上加以预测。一种设计巧妙的连锁装置，使得当放射性样品中的某个原子发生衰变时，它触发的信号能使一把预先定好位置的榔头落下，打破小瓶使有毒气体逸出，从而把猫杀死。按照常识，猫是非死即活；但是按照量子力学原则，由箱子和其中一切物体所组成的系统，是由一个波函数来描述的。为使问题简化，让我们假设这猫只能存在于两个量子力学态——活或者死，系统的波函数中，就包含着这两种可能的、但相互排斥的观测结果的组合。因而这猫在同一个时刻是既活又死。如果我们这位老好人薛定谔教授不去打开箱盖看这猫，他自己的方程就表示，这猫的时间演化在数学上，可以用这两种状态的组合来描写，而这种组合在物理上（以及生理学上）是说不通的。正像在测量做出之前电子既不是波也不是粒子一样，这猫在教授最终决定去窥视它以前，既不是活着也不是死了。

薛定谔就是用他发明的这套“地狱般的装置”去抨击量子力学的非决定性，他把这非决定性从放射性衰变的微观尺度，转移到了死猫的宏观尺度。观测的作用不仅明显地在现象中注入了一种主观因素——某个人必须打开箱子去看这只猫，而且它也迫使猫不可逆地接受这两种可能性之一——要么玻璃瓶完好无损、猫也安然无恙，要么瓶子被打碎从而猫死去。

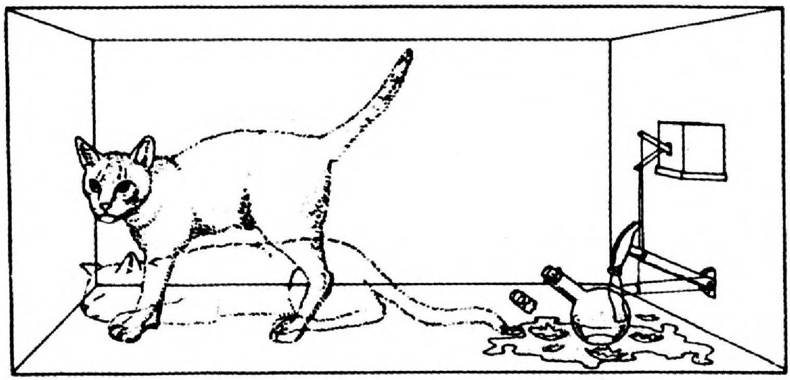


图12　薛定谔的猫和放置它的那个恶魔般的装置，箱壁假设是不透明的。如果箱壁是透明的，则按照量子基诺佯谬，放射性物质就永远不会衰变因而猫也永远不会死去。[录自《量子力学中的多世界论》（B.S.dewitt，N.Graham编辑）第156页。]

薛定谔的猫生动地把测量问题摆到我们的面前。看来我们得要相信，系统的状态被观测本身改变了。然而这显然又太离奇。如爱因斯坦所说："我不可能想象，只是由于看了它一下，一只老鼠就会使宇宙发生剧烈的改变。"然而，即使是量子力学的先驱们，也没有把这当做必定是一件坏事。正像玻尔在与瑞士理论物理学家泡利（Wolfgang Pauli）的一次交谈中，对泡利说道，泡利的一个想法"是狂想，但狂得还不够"。

## 魏格纳的朋友

有两种简单通俗的办法，试图去对付爱因斯坦的异议。其中之一根据的是这样一种观念，即对量子系统进行测量的，不是猫或者老鼠，而是富于意识的人。这种想法的提出者中间，包括纽曼和魏格纳（Eugene　Wigner），他们认为需要具有意识的观测者去"看"，从而影响波函数的坍缩；根据他们的看法，猫不具备观测者的资格以使波函数坍缩到生和死的真实情况，因为它并没有足够聪颖可以区分出这两种状态。按照这种观点，精神和物质是完全不同的观念，而且量子

论只对后者有用。只有具有意识能力的精神才可以触动波函数。

这样就产生了一个推断，薛定谔的那可怜的猫，并不知道它自己是活着还是死了（纵然大多数宠爱猫的主人都声言，他们的宠物有和人类相近的智慧）。然而，动物是不是具有意识，这是一个极有争论的问题，起码这也是因为意识这个词本身就很难科学地下定义。它通常用来指人类那样的智慧，虽然这样一个泾渭分明的概念在这里很可能无助于事。

现在我们再来看“魏格纳的朋友”，他是这位诺贝尔奖获得者想象中的熟人。他的这位朋友戴着防毒面具，吃力不讨好地和猫一起关在箱子里。当他睁开眼睛时，猫的波函数就坍缩了。在薛定谔往里面窥视之前，这位朋友可以回答在箱子里面的感受。他可以用习惯的语言来报告发生了什么事——在箱盖未打开之前，他根本没有被卷入所有可能的实验结果的量子迭加状态。这样，按照魏格纳的说法，当人类的智能也包括到所研究的系统之中时，通常的那种量子描述就不能再用了。根据这种基于意识的量子理论解释，时间的流逝只不过是一种心理效果，它相应于不断地触动波函数。

人们可能会问道，魏格纳的朋友要喝多少升啤酒以后才失去知觉，从而失去触动波函数的能力。看来，我们现在很有回到一个极其主观的世界观的危险。正像雷（Alastair Rae）在他的《量子物理学：幻觉还是现实？》一书中所写道的：“自从现代科学四五百年前开创以来，科学思想似乎已经把人类和意识从世界中心远远移开了。宇宙中越来越多的事物，变得可以用力学和客观的术语来解释，即使人类本身，生物学家和行为科学家正在用科学的方法加以了解。而现在我们却发现，物理学——以前被认为是所有科学中最客观的科学，正在重新需要人的灵魂，并把它放置在我们对于宇宙理解的中心！”科学所积累的证据是如此之多，有利于一个独立于意识之外的现实存在，这使我们很难重视一个基于意识的处理量子理论的办法。确实，也许有一天，意识本身都会用物理学术语加以解释。

# 平行的宇宙

为了“解决”量子力学中的测量问题，埃弗雪特（Hugh Everett III）1957年提出了另一个大胆的办法。他这个后来被科学幻想作家经常采用的想法，萌生于他在普林斯顿大学做博士研究期间，当时热心指导他的是惠勒教授。让我们回到杨氏狭缝实验并来考虑这两难推理，即光子到底是从两条狭缝中哪一条中经过。波函数中包含了光子从两条狭缝中经过的所有可能的结果，然而在最基本的层次上，对光子所经过的狭缝是要有一个选择。哥本哈根解释说，这就是按照几率法则对不可逆坍缩的选择。

然而，埃弗雪特的解释说，电子不是选择狭缝，而是选择宇宙。在选择其中一条狭缝而不是另一条时，宇宙就一分为二。这条被选择的狭缝决定于我们处在哪个宇宙。此后这两个宇宙就完全分开了，并且越分越多，每做一次测量，宇宙就分裂一次。埃弗雪特的想法关键是，宇宙自身是由一个波函数描述的，这个波函数中包含任何实验结果的组成部分。他的解释中有一个异乎寻常的含义——独立存在着无数“平行的”宇宙，每一个宇宙都像我们的宇宙那样真实。你做的最荒唐的梦，也许就发生在另一个世界。一个被定义在某个宇宙中的观测者，他所做的每次测量，都使这整个宇宙萌发出无数多个新宇宙（即“多重世界”），每一个新宇宙代表一个不同的、可能的观测结果（例如一活着的或死了的猫）。没有波函数的坍缩发生，只有新分支出的宇宙的无穷尽的增殖和萌发：不需要有一个宇宙之外的观测者。埃弗雪特的论文，与9世纪伊斯兰经院神学卡拉姆派的教义很有些相似：按照这个教义，随着每一个事件的出现，世界得以再生。

宇宙学家德威特（Bryce DeWitt）这样描述过他第一次听到埃弗雪特主张时所受到的震惊：“我仍然清晰地记得，当我第一次遇到多重世界概念时所受到的震动。100个略有缺陷的自我拷贝，都在不停地分裂成进一步的拷贝，而最后面目全非。这个想法是很难符合常识

的。这是一种彻头彻尾的精神分裂症……”

许多宇宙学家偏爱埃弗雪特的多重世界量子力学解释，因为它不再明显地需要有一个外部观测者。这一点是重要的，因为如果采用魏格纳的解释，能够使传统的宇宙波函数坍缩的唯一观测者，必定是上帝。相比之下，埃弗雪特的主张看来是“在假设方面代价低廉，而在宇宙方面代价高昂”。有意思的是牛津大学的道奇（David Deutsch），他认为他所提出的量子计算机，一旦建造完毕，将可以从实验上验证埃弗雪特猜想的正确性。这种计算机具有许多新奇的性质，包括一种量子并行机制，使许多计算可以同时进行，而且计算速度也比通常的计算机快得多。道奇声称，这种奇迹般的快速计算，将需要把不同的计算部分放在一些平行的宇宙中进行。他的说法引起了很多争论，因为没有科学证据支持埃弗雪特的猜想，并且即使这样一种计算机造出来，也未必能证明多重世界一定就比其他解释优越。除此以外，多重世界的解释具有很多技术性的困难，特别是，它没有说明究竟是什么特殊原因，使得测量过程会导致宇宙的分裂。总而言之，它给人的印象是：杀鸡用牛刀。

## 读者存照

量子力学基于意识的解释和多重世界的解释，它们的牵强附会，反映着在现有的量子论的框架中，测量问题所引起的困难。基本问题被搞得含混不清，原因是人们广泛地、但错误地认为，传统的哥本哈根解释需要“观测者”的存在——这“观测者”被假定为人类，从而使主观因素乘虚而入。但是，其实并不需要有“意识”的生物参与：只要放上一套测量仪器，记录实验的结果就足够了。例如，这套仪器可以是由一台计算机，一个荧光屏，一张摄影感光乳胶片或是一个气泡室组成。人的意识与之无关。波函数的坍缩是不可逆的并且是完全客观的。

# EPR佯谬

如果否定基于意识的解释和平行宇宙的解释，我们仍然面临着一个主要困难，即如何说明波函数的坍缩。这种坍缩看来违背了因果律常识——即果起于因，因果事件在空间上必须足够接近，以使得它们之间发生因果性联系。

考虑一个量子法则支配的系统，比如一个朝着荧光屏飞去的电子。它随时间的演化由一个量子力学波函数来描述，其演化规律遵从薛定谔方程。当在屏幕上做测量，以确定电子的位置时，波函数就坍缩了。虽然测量是在局部地点进行的，但是波函数的坍缩却改变了电子在此瞬间在空间所有地方的物理状态——要记得，波函数中包含了电子在空间和时间中的每一种可能性，并不只是波函数坍缩时的那一种可能性。

这种非局部的坍缩意味着违背了因果律。有一些时空区域，它们不能与屏上观测到电子的那个点有因果联系，即使信号以光速传播也不行。然而它们却可以在瞬间，和在这个地点所做的测量发生关联！电子发出的作用瞬时间达到宇宙的尽头，这听起来就像一位巫术师，能用针刺蜡像而伤害几千米以外的人一样，纯属天方夜谭。这一独特性质出现的原因，是由于波函数实际上是一种弥漫于整个空间的抽象场。一个在任何一个给定地点所做的观测，会引起一个潜在的状态（所有可能的结果）坍缩到一个现实的状态（观测到的结果），这个现实状态在宇宙中所有各处，并在同一时刻被固定下来。

这种离奇的非因果性的相关性，是1935年首次由爱因斯坦和俄国的波多尔斯基（Boris Podolsky），以及罗森（Nathan Rosen）在一篇论文中强调指出的，后面这两个人后来都成了普林斯顿高级研究所的成员。爱因斯坦本人是在1932年获得这个研究所的一个终生职位的，这使他逃脱了战前德国日益高涨的反犹太浪潮。这篇文章发表的时候他已经在普林斯顿定居，他的狭义相对论问世也差不多30年了。

但是量子论仍然使他苦恼不安，一是量子论中的内禀随机性，二是现实看来是由观测者而“创生”的。对爱因斯坦来说，这种说法像是把人又推回到500年前宇宙中心的地位，而哥白尼当时就是把人从这个地位拉了下来。因而，爱因斯坦极力要把量子力学的这种观念推翻。

爱因斯坦和他的同事所构想的“思维实验”——取他们三个人名字的第一个字母，因而叫做EPR佯谬，是设想两个有旋转的粒子，它们相互作用并分离到无限远的距离。我们从它们相互作用的方式知道，每一个粒子具有一个等量但符号相反的“自旋”值：如果一个的自旋是“向上”，则另一个的就“向下”。这样，粒子A的自旋，就可以从粒子B的自旋推断出来。但是在量子力学中，两个粒子的自旋都处在一种不定的状态——“向上并且向下”——直到测量做出为止。测量使得粒子A的自旋在测量时刻成为“向上”或者“向下”，但是由于量子相关性，粒子B必须也立刻被迫接受一个确定的自旋，即与A的自旋相反。这一结果确保成立，不论粒子B与A的距离已经是多少个光年。我们不需要去测量B的自旋，因为一旦A的自旋测出，B的自旋也就知道了。这种遥远距离上鬼使神差般的作用看上去意味着两个粒子之间的联系，是由一种传播得比光还快的物理作用来进行的。

由这一分析可以看到，EPR佯谬的论点是，量子论不能给出一个现实存在的客观描述，并且量子论也是不完备的，因为它包含了如此非物理的、“非局部的”相关性。从那以后，很多研究都在探寻量子描述的某种更深一层的基础——即一种“隐变量理论”，它或许会以一种决定性方式说明这种相关性，并且可能进一步接受最后仲裁——即实验验证。这样的验证已经按照一个实际的EPR实验的路子搞了起来，它的结果会由于隐变量所支配或者是量子论所支配而有所不同。

对爱因斯坦和他的追随者来说不幸的是，量子论还是赢了。1982年，阿斯佩克特（Alain　Aspect）和他的同事在巴黎理论和应用光学研究所证实，看来确实有一种超光速的联系存在于遥远的时空区域之间。两个在宇宙中远远分离的粒子，可以以某种方式组成一个单一的

物理整体。这样看来，量子不确定性所暗示的那种疯癫独特的宇宙确是存在着：上帝的确是在把宇宙当做骰子玩耍。并且我们必须得出结论说，爱因斯坦所想象的、一个完全用科学描述的决定性实在，只是一个无法捉摸的幻想，它来自我们对世界的“常识”看法。

## 当时间停滞了

波函数坍缩的不可逆性，为时间之箭的客观存在提供了强有力的证据。然而，客观的波函数坍缩是以其特殊形式出现的，因而它不能对瞬息即逝的坍缩给出我们想要的完满解释。这一点可以从这个理论的一个显著特点看到，它表明，当一个量子系统被连续地观测的时候，时间实际上停滞了。让我们来考虑一个不稳定的原子核经历放射性衰变的情况，就像在猫佯谬中发生的那样。在量子论中，测量往往被理想化，是在某个固定的时刻进行的。但现在我们设想，原子核是在被连续地测量，目的是确定它究竟在哪个时刻发生衰变。

20世纪70年代，奥斯汀得克萨斯大学的米斯拉（Baidy-anath Misra）和苏达山（George　Sudarshan）发现，在这样的条件下，原子核就不会发生衰变。这就像是把“盯着的水壶总是不开”这句谚语用到量子力学中来。连续的测量迫使原子呆在不衰变的状态，使它不能转换成衰变生成物。为了引起人们对他们的发现的注意，米斯拉和苏达山把这种现象称为“量子基诺佯谬”，因为它与基诺佯谬具有某些相似性。基诺佯谬是讲在空中飞的一支箭，但是根据他的分析，这支箭不可能在运动。然而，这量子力学中的比拟，并不能算是一个佯谬，它只是正统量子理论的一个推论罢了。

1989年在美国科罗拉多州波尔德的国家标准和技术研究所，有一个研究小组做了一次非常精妙的实验，实验结果支持了量子基诺效应。研究人员对放置在磁场中的5000个荷电铍原子进行了观测。这些原子开始时处于同一能级，由于一个射电频率的电磁场的照射，它们

可以在0.256秒内“煮开”，即激发到高能级。只要在这当中不进行任何测量，所有的原子在照射之后都会位于高能级。然而，当用激光在这期间的某个瞬间进行探测时，研究人员发现，他们探测的次数越多，能达到高能级的原子数目就越少：当每隔千分之四秒进行一次探测时，就没有一个原子能够被激发。这样看来，一个被盯着的量子水壶是不会煮开的。

连续测量的概念，和分立测量一样，都是理想化的。任何实际的测量都只是持续一段有限的时间，这样，量子跃迁最终总会发生。这两种极端情况的结果显然都有些过于绝对：或者是某件事发生了，或者是什么事都没有发生。让我们再研究一下猫佯谬中那个凶神恶煞般的装置，但是把箱壁从不透明的换成透明的。这样我们就可以在所有时间中观察，到底箱子里面是怎么一回事。这时候量子基诺佯谬会说，只要我们保持观察（当然“我们”也可以换成为无生命的测量装置），猫就一定是活着的。看来，我们所谓的“伟大理论”确实是有问题。

薛定谔的猫、魏格纳的朋友以及其他佯谬，都突出地表明波函数的坍缩需要某种解释。在我们看来，出现这些困难的原因，是由于现有的量子理论是时间可逆的，而测量过程是内禀不可逆的。基于意识的解释和多重世界的解释，显然都是由来已久的削足适履之术。这表明，今天的量子理论还是不完备的。测量所造成的问题，只有借助于结合时间箭头的理论才能解决。我们将在第八章中讨论这一点。

## 物质和反物质

现在，我们还是继续对时间之箭的探索。一个对寻找时间之箭具有魅力的地方，是在亚原子粒子的奇异世界。我们所依据的基本原理，出自英国数学物理学家狄拉克（Paul Adrien Maurice Dirac），他在20世纪30年代量子论和狭义相对论相结合的过程中，起了关键性

作用。开始的时候，狄拉克学的是电气工程师，但是他很快发现，他的爱好偏向于物理学。他有善做智力体操的令人敬畏的名声，以及怕羞、孤独的禀性。有一次，当他在多伦多大学作完演讲之后，听众中有一位加拿大教授问道：“狄拉克博士，我还没有搞懂，你是怎么样推出黑板左上角的那个公式的。”在一阵长时间沉默之后，并且也是在会议主持者的提醒之下，狄拉克这才回答说：“他说的不是一个疑问句，而是一个陈述句。”

薛定谔从一开始就认识到，他自己的方程有局限性——不满足爱因斯坦狭义相对论的要求（特别是，它不是洛伦兹不变的），而且它也不能说明原子光谱的精细结构。狄拉克针对这一问题，在1928年得出了一个相对论性的方程，它表明电子必须绕自己的轴自旋。除此之外，狄拉克方程的数学特点使他提出，应当存在正电子，这是“反物质”的一个基元。正电子的质量和电子的一样，但是电性相反。正负电子的碰撞会使得两者同时湮灭，而产生辐射爆发。安德森（Carl Anderson）1932年所做的一个实验，证实了正电子的存在，因而使反物质成为现实。这个发现从根本上改变了基本粒子物理的基础概念。在此之前，物理学家们对古希腊物质不变的观念深信不疑。这个实验之后，人们认识到物质可以任意产生和消灭。

根据狄拉克的数学描述，正电子被解释为如同它的反粒子——电子——在时间上向后退行。费曼甚至于提出整个宇宙中只有一个电子，它在时间上时而向前时而向后，轨道运动是如此复杂，使我们以为在任何瞬间都看到了大量单独的电子。像我们在第三章谈过的戈德尔的时间旅行一样，费曼的“一个电子”理论需要不同的时间方向同时存在，这导致与因果性观念的直接冲突，并且不可避免地，它产生的问题比它解决的要多。

在物质、反物质、空间对称性和时间的两个方向之间，存在着一个错综复杂的关系。它出现在一个奇特的CPT定理之中，这个定理是物理学微观定律数学形式的一个结果。CPT定理来源于一些定律的对

称性，把任何过程中的粒子换成反粒子、把该过程换成它的镜像（即如同在一面镜子里看到的该过程）以及把时间倒转这三个变换同时操作的情况下，这些定律保持不变。这个定理是吕德斯（G.Lü iders）于1954年，泡利于1955年分别得出的。它的名字起因于三种抽象操作，它们的共同结果使得对称性得以保持：

◆C电荷共轭，它把物质转换成反物质；

◆P空间反演，它把空间坐标转换成它的镜像；

◆T时间反演，它把时间方向倒转。

CPT三重操作对于一个过程作用的结果，产生了另一个同样容许的过程，它也由同一理论框架所描述。不太严格地说，CPT定理断言，物理规律预言了在一种“泛镜像”世界中的等量但相反的事件。它同时可以告诉我们，时间对称性是如何可能被破坏从而产生出时间箭头。

CPT形式下的对称性可以用来演绎物理学。这样我们会发现，量子定律对于一个飞过球场的板球的描述，与对于这个板球的反物质的镜像在时间上倒退，最后飞回到投球手手里的描述是一样的。CPT定理最重要的一点是，如果所讨论的过程是CP对称的，则它必须也具有时间可逆的对称性（T）。这在本质上类似于，我们可以从一套餐具中餐叉的数目，推断出餐刀的数目。CP对称性还意味着，如果有一只反物质的左手戴着一块反物质的表，则一定会有一只正常物质的右手戴着一块正常物质的表；这样T对称性就表明，每一只表在时间上既可以向前走也可以向后走。相反的，如果CP对称性受到破坏——也就是反物质的左手不再有相应的正常物质的右手——则T对称性也要受到破坏：左手上戴的表在时间上就只能向前而不能向后。这看来就按归并派的想法，对于时间之箭在理论上给出了一个基本原理，它显然与前面谈到的测量问题完全无关。然而，几乎所有已知的“基本粒

子”的相互作用都满足严格的CP对称性，因此它们对时间进行的方向并没有特别的要求，因为它们是时间对称的。

## 长寿K介子奇案

在微观“基本”粒了物理世界中，仅仅有一个奇特的现象，被认为是破坏了时间两个方向之间的对称性。这出现在一种不稳定的长寿K介子（$K^{\circ}$）的某些衰变过程中。K介子是在美国的布鲁克哈文，由克里斯坦森（J.H.Christenson）、克罗宁（J.W.Cronin）、菲奇（V.L.Fitch）和特雷（R.Turlay）发现的。这一发现使克罗宁和菲奇获得了1980年的诺贝尔物理学奖。在大多数衰变中，K介子生成一个负的π介子、一个正电子和一个中微子，这一个过程的CP对称性被证明是保持不变的。然而K介子同样也可以衰变为（大约在10亿次衰变中有一次）一个正的π介子、一个电子和一个反中微子，这时候CP对称性就会受到破坏。按照CPT定理，T对称性在这罕见的过程中也就同样受到破坏：时间可逆的事件被禁止，过程成为不可逆的，时间箭头便显现出来。

在我们接受有时间意识的微观现象之前，哪怕这样的事例只有一个，我们首先应当探究CPT定理本身的正确性。CPT定理的证明是根据一些假定，但如果引力也包括在内的话，这些假定便不能满足了。然而，如果在这种罕见的过程中发生CPT对称性的整体性破坏（还没有观测到这种情况），这种破坏要比观测到的CP破坏还要小；因此T破坏一定发生了，虽然是间接地发生。彭罗斯评论说：“这几乎是完全隐藏着的时间不对称，它微小的作用看来在K介子衰变中是真实地存在着。很难相信，大自然不是在通过这一精美漂亮的实验结果，来试图给我们‘指点迷津’。”

但是，对这些内禀的时间不对称的相互作用，目前还没有令人满意的基本解释。因而还不能确定，是否长寿K介子的奇怪衰变将使我

们对时间之箭的搜寻转向，或者如彭罗斯认为的那样，它将成为解释时间之箭的关键。这是因为，在本书以后要描述的一些熟悉的不可逆过程中，它并没有起到什么作用。

## 沸腾的真空

如我们已经看到的，海森伯测不准原理影响到时间的测量。在一个给定的时间间隔之内，我们所能够测量的能量的精度有一个限制。精确测定一个原子处在一个特定量子态时的能量，是要以处于这个态上的时间之相当大的不确定性作为代价。

下一节中我们会看到，一些宇宙学家相信，可以从这种不确定性中生出一个完整的宇宙。不确定性关系使得能量不能无中生有这个观念可以被违反。在经典物理学中，能量既不能创生也不能消灭，而是严格守恒的，只是从一种形式转换到另一种形式。例如，汽油中的化学能转变为热和汽车的运动。对于所有的初始能量，可以按这种方式做出一份能量平衡表。但是，如果时间间隔取得过小，能量守恒就会由于海森伯不确定性原理而受到破坏。

海森伯原理中的能量和时间的关系表明，所考虑的时间间隔越短，则能量的不确定性就越大。这使得能量守恒在非常短的时间间隔内不再成立：由于随机的量子涨落，能量可以从虚无中得到。这样的事件甚至可以在真空中发生，而按照经典的看法，真空是一无所有的。这样，量子论就给出了一个完全不同的真空概念。由于不确定性原理，真空实际上沸腾着活力。

现代的真空概念主要是狄拉克建立起来的。他认识到，如果要正确地描述物质吸收和发射光子的方式，麦克斯韦电磁场就必须用量子术语来描述。他推广了麦克斯韦的数学模式，把电磁场描述为数目巨大的振子的集合，每一个振子的能级都是量子化的，像原子中电子的

能级那样。但是现在，由于不确定性原理，每一个振子的能量不能低于一个固定的最小值——即零点能，这就使得即使是真空也总是沸腾着活力：在空间所有各处，真空场的能量永无终止地在发生涨落。足够大的能量涨落可以使得粒子——反粒子对——例如电子——正电子对——在瞬息间生成，而且能量涨落越大，粒子对生成得就越迅速。这些真空量子场涨落具有相当重要的物理意义。例如，光子不断地产生和湮灭，可以触发吸收了能量的原子，使其自发地发出像光，即辐射。事实上，真空的涨落在某种程度上使虚无缥缈的以太重新受到注意，以太当时是被爱因斯坦当做一个多余的累赘，在1905年丢弃了。如牛津大学的史密斯（Christopher Llewellyn Smith）所说："今天，我们对于真空一点都不懂。"

狄拉克把电磁场量子化以后解决了一些问题，然而却引起了更多的问题。简言之，困难的出现是由于场所能够携带的能量没有限制，这就导致理论中常常出现无穷大。正像广义相对论中奇点的情况一样，这种失控的行为在数学上是很讨厌的，并且它暗示着理论框架中什么地方有了问题。但是，在量子场论中，一些处理问题单刀直入的理论家，例如戴森（Freeman Dyson）、费曼、史温格（Julian Schwinger）和朝永振一郎（Sinitiro Tomonaga）发现了一种称为"重整化"的办法，可以用来克服这种发散困难。这种办法使得无穷大与另外的无穷大相抵消，因而被巧妙地吸收掉了。结果的形式给出了富于意义并且常常是非常成功的预言。但是，围绕重整化方法合理性的争论一直在继续；狄拉克本人就认为，无穷大问题是理论本身确实具有基本缺陷的征兆。甚至霍金都承认，重整化"在数学上是值得怀疑的"。

量子场论的主要缺点之一，是它不能处理引力。这时候产生的无穷大，即使借助于重整化技巧也无法消除。然而，宇宙学家们普遍认为，一个成功的量子引力理论，会解决我们在第三章中遇到的广义相对论的奇点问题。最近，出现了一种解决这个困难的新办法，即所谓

弦理论。可以把弦理论看做是某种“高维”的场论，它的最小单元不再被看成是点，而是某种具有有限尺度的东西，即开放的或闭合的弦。这一理论的主要倡导者中间，有伦敦大学玛丽皇后学院的格林（Michael Green）和美国加州理工学院的施瓦兹（John Schwartz）。自从弦理论一问世，很多人都对它抱有乐观的希望，认为它将能够处理并统一包括引力在内的粒子间的基本相互作用，不会出现棘手的无法控制的无穷大。正因为如此，一些持高度乐观态度的物理学家，其中包括霍金，认为随着弦理论的出现，理论物理学的终点已经在望。

虽然从数学上讲，弦理论具有无可否认的美学上的吸引力，但在科学上还没有使人非相信不可的理由，能把它作为一种万能的灵丹妙药。用巴罗的话来说，“还没有实验上的事实证明它正确与否。在今后一些年里……一定会有这样的事实出现。只有在那以后我们才会知道，这一独特的处方是一种包罗万象的理论，还是一文不值。”尽管弦理论自己宣称可以解决许多问题，但是由于它的时间对称结构，它看来不大可能对时间的本质给出任何新的见解，特别是关于时间的方向，或者是与此有关的、然而常常被忽略的测量问题。

## 宇宙免费午餐

我们回忆一下，由于存在无法避免的、棘手的奇点问题，用西亚玛的话来说，广义相对论“本身包含了自我毁灭的种子”。虽然如我们刚刚提到过的，至今还没有一个可以令人满意的量子引力理论，但是在过去的10年左右的时间里，现代宇宙学家们已经开始思索，量子论如何可以使爱因斯坦相对论中令人不快之处得以改善。

量子论和相对论的特殊融合所产生的量子引力理论，导致了一些关于宇宙和时间本身创生的极有趣的想法。例如，沸腾的真空就启发出一个宇宙如何诞生的模型。如果引力是量子化的，则必然有引力场

的随机涨落，这就给出了一个“无中生有”的“机制”。美国麻省理工学院的盖斯（Alan Guth）是当代宇宙学家中一位主要人物，他说道：“我常听人讲，没有免费午餐一类的好事让你遇到。但是现在看来，宇宙本身就是一份免费午餐。”这个想法根据的是真空沸腾的概念，即由于海森伯不确定性原理，真空的能量发生无规则的涨落，如此就产生出宇宙——这是美国物理学家特雷恩（Edward Tryon）于1973年首次提出的，但是在开头5年里他的想法并没有引起响应。

从1978年开始，比利时布鲁塞尔自由大学的一个宇宙学小组，提出了一系列引人入胜的建议。按照他们的“圣经”，宇宙在原初时刻是“虚无”的真空，时空也是平直的。这和令人不快的大爆炸奇点完全不同，在大爆炸奇点，宇宙中的一切在初始时刻是被压缩到一个点的。从爱因斯坦著名的质量和能量之间的关系，能量中的量子涨落将会产生出等量的粒子质量。这样一来，经过粒子间的相互吸引，就会使初始平直的时空变成弯曲的，如我们在第三章中已经看到的那样。这一过程并不破坏能量守恒，因为粒子质量吸收的正能量，被引力吸引的负能量所补偿了。这个小组的宇宙学家们认为，虚无的真空状态相对于物质创生来说是不稳定的：像是一个连锁过程，一旦开始，就会导致宇宙中所有的物质和能量自发地创生。

另一些宇宙学家随之也提出了其他的模型，但每一个都只是对上述模型作某种修正或改进。威廉金（Alex Vilenkin）以及稍后的霍金和哈特勒（Jim Hartle），试图给出空间和时间从绝对一无所有中创生——这种状态，甚至于不是沸腾着物质的量子力学产生和湮灭的“虚无”真空。霍金和哈特勒想把时间开始的奇点即大爆炸敷衍过去，他们对整个宇宙给出了一个波函数，然后计算从确确实实的一无所有中产生出某种东西的几率。

霍金的方法中，部分地采用了费曼的“经历求和法”和所谓的“虚时间”。按照霍金的说法，虚时间“听上去像是科学幻想，但实际上是一个有确切定义的数学概念”。虚数自乘得到的结果是负数。而对于大

多数人熟悉的实数来说，＋2和-2的平方一定都是＋4。虚时间像爱因斯坦相对论中的普通时间一样，是没有方向的，采用虚时间以后，空间和时间的数学形式之间就完全没有区别了。在闭合的宇宙的情况下（换句话说，也就是有大坍缩点的宇宙），这种混在一起的时空可以存在于非常早期的阶段，它在尺度上是有限的，然而没有边界或令人不安的奇点。时间在宇宙的极早期会没有定义，而宇宙本身看上去很像地球的表面，但是要多出两个维度。这种自我封闭的宇宙被认为是不需要初始条件的，因而爱因斯坦理论的奇点困难也就得以克服。

虚时间其实只不过是一种数学技巧，它并没有告诉我们关于实在世界的任何新东西。然而，霍金的看法是，“如果你也像我一样采取实证主义的立场，那么任何有关现实实在的问题都没有意义。我们所能问的只是，在描述我们观测到的东西的数学模式中，虚时间是否有用。答案是肯定的。的确，我们甚至可以采取极端的立场，认为虚时间其实是基本的概念，而数学模型必须用它来表述。普通的时间反而是一种派生的概念，我们发明了它，是把它作为数学模型的一部分，用以描述我们对宇宙的主观印象。”

对于这些讨论“无中生有”的纯理论模型，保留一定的怀疑态度还是有益的。对于它们，有许多技术上的反对理由，也有许多原则性问题。作为后者的一个例子，如果“无”确实意味着“一无所有”，则看来我们并无任何权利，把我们的科学定律运用到创生行为之前。这样我们就会仍然疑惑不解，是什么把宇宙带到现实存在中来。

但是，关于时间之箭又是如何呢？对于这个基本问题，所有这些模型都保持沉默。因为量子力学（除去令人棘手的测量问题）和相对论，两者都是时间上严格可逆的。在第三章中已经提到过一个非常令人兴奋的可能性，这就是彭罗斯的建议，即迄今为止还不了解的“真正的”量子引力，应当是时间不对称的。也就是说，它应当明显地含有时间箭头。我们将在第五章谈宇宙学和热力学的关系时，再讨论彭罗斯的理由。这样的一个根本性建议，可能会最终导致物理学的一场

新革命，而我们现有的视野还不足以使我们对此做好准备。

在这一章中，我们概括地评述了量子论令人困惑的课题。我们的结论是，它充其量也只是一个不完备的理论。主要的困难是围绕着测量行为，因为在测量过程中，包括所有可能结果的波函数坍缩到单一的现实结果——例如著名的薛定谔思维实验中那只猫的死亡。由于测量过程内在的不可逆性，它不可能在现有的可逆性的理论框架中得到体现。

下一章我们要讨论热力学，它是一种包含时间之箭的不可逆过程的科学理论；以及统计力学，它试图在微观量子世界和宏观热力学世界之间架起一座桥梁。这样，我们也许会对现有的量子描述中的缺陷有更多的见解。

# 第五章　时间之箭：热力学

对物理现实的图像，尤其是对时间的属性，我们来一次大变革的时候到了。这变革也许甚过令日的相对论和量子论。

——彭罗斯

《皇帝新脑》

当我们从量子力学的微观世界转移到日常生活的宏观世界来，时间之箭就变得比较清楚了。这就是热力学的领域，它是一个威力巨大的理论，其中时间的流逝方式与萦绕在诗人和小说家脑际的想象是一样的。格雷夫斯（Robert Grayes）有一次生动地描写道，时间就是“计数脉搏，计数缓慢的心脏跳动，在缓慢的心脏跳动中流血、从而走向时间的死亡”。热力学所做的与此相同。它揭示了同样不可逆的过程中瞬息即逝的现实，从我们的青春逝去到眼泪风干。这些不可逆过程使人类的存在既富于深刻哲理又富于情趣。当然，并不是它的所有应用都是这么富有象征性。热力学还解释蒸汽机如何工作以及为什么茶会变凉这一类问题。

热力学把时间与有序性和无序性（随机性）这样的概念联系起来。时间的流动变得显而易见，这是因为在任何孤立系统中都有一种毫不留情的倾向，使得有序程度降低而无序程度增加。如果往红茶里面加一点牛奶，奶分子就会与茶分子混合在一起并且扩散。最后，奶和茶的分布会完全一样，茶显出特有的浓褐色；当混合过程完成以后，不会再有进一步的变化发生。在茶所达到的最后状态中，分子的无序性——或者用热力学术语准确地说，熵——达到了一个极大值。

这是一种平衡状态，奶分子和茶分子在混合物中所有各处都是均匀的，不再具有任何进一步混合的能力。我们从来没有见过相反的过程，即均匀的褐色液体自发地分开成为白色的奶和红色的茶。因为要使这样的过程发生，我们就必须让时间开倒车。

时间之箭在所谓的“热力学第二定律”中明显地表现出来。这个定律说，所有的物理过程都是不可逆的，因为一部分能量总是要作为热而散失掉。作家斯诺（C.P.Snow）认为，对于任何受到过良好教育的人，热力学第二定律都应该成为他所应具备知识的一部分。他描写过一些受过高等教育的人，“他们津津乐道地形容科学家们对文学的无知。有一两次我被激怒了就问他们，你们中有多少人可以讲一讲热力学第二定律。反应是漠然的，而且答案也是否定的。然而我的问题实际上等价于问一个科学家：你读过一篇莎士比亚的作品吗？”

不幸的是，正像文学批评家们为莎士比亚的戏剧争论不休一样，科学家们也为第二定律的意义发生激烈的争执。热力学的基本命题相当含糊，从而导致了多种多样的观点。这种情况也反映了通常对科学家的印象之荒唐：一般人总认为科学家是一伙穿着白外套、冷面孔、思想一致的人。事实上，科学家之间观点上的差别之大，常常可以表现为个人之间的冷嘲热讽和激烈的争吵。美国哲学家赫尔（David Hull）最近指出，物理学家们可以轻而易举地罗列出20种或者更多的第二定律的不同表述形式。赫尔写道，给局外人强烈印象的主要是，每一个科学家都坚持说他自己的观点是正确的。同时，像热力学本身一样，使人伤脑筋的时间问题，经常产生的更多的是争论的热度，而不是希望之光。

当我们讨论热力学的不同处理方法时，一些观点的相互冲突会变得很明显。我们会看到一种根据热力学平衡的描述，它是一种特殊情形，此时作为热力学理论核心的一切变化和流动都被抑制，而这种情形又是大多数科学家在热力学课题中唯一了解的部分。我们将要表明，平衡态热力学是一条死胡同。某些人错误地认为，热力学排除了

有序结构的自然出现，而且意味着宇宙的整个进程是直接走向无序的混沌。我们会看到这也是不正确的，实际上热力学理论正把握着有序生命产生的关键。但是对我们眼前的目的来说，最基本的问题还是，如何使热力学和分子力学的微观理论相和谐，因为上面我们已看到，这种微观理论是时间对称的。科学思想的这两个至关重要部分之间表面上的抵触，一些人称为不可逆性佯谬（另一些人称为可逆性佯谬）。它已经使得一些科学家认为时间的热力学箭头在我们的精神之外，不具有现实性，是一种纯主观的概念。

## 热力学的诞生

热力学理论是随着19世纪早期，英国工业革命中蒸汽动力的出现而形成的。第一台实用的蒸汽发动机是1782年瓦特（James Watt）建造的，他以前在格拉斯哥大学从事科学仪器制造工作。蒸汽机烧煤把水加热，因而产生蒸汽压力推动活塞或转动涡轮桨叶。但是为了计算一台发动机的最大效率，就必须了解这台机械幕后的全部理论。这一学科就是热力学，它来源于希腊语，意思是热的运动。

当蒸汽机运转时，能量转化为所有组成部分的分子运动，这一过程是极端复杂的。但是热力学并不涉及原子和分子（我们回忆一下，当时原子的概念还没有得到普遍认可），它把注意力直接集中在一些与感觉有关的“宏观”量上面，像体积、温度和压力。在早期热力学家中一个杰出的人物是法国工程师卡诺（Sadi Carnot）。他是法兰西第一共和国一位领导人的长子，后来在一场流行性霍乱中去世，终年只有36岁。然而在去世前10年，他就已经对理想热机如何工作给出了一个透彻的热力学分析。他的这种理想热机是完全可逆的，没有不可逆的热损失。他说明了热机的效率如何决定于热量从热物体流到冷物体这一事实。在一台蒸汽机里，这就是说，热量从蒸汽形成的热汽室，流到蒸汽凝结的冷汽室。卡诺说，冷热物体之间的温度差异，就决定了热机工作的好坏。有重要意义的是，即使这样一种完美的热机，它

的效率也绝不可能达到百分之百。

用实际用语来说，蒸汽机的工作就是把热转化为功，这里功的意义就是一种更有用、更有组织的能量。热和功的等价，是曼彻斯特一个酿酒世家的儿子焦耳证明的。他把一台蒸汽机的工作倒过来观察，也就是用功来产生热（用桨来搅动水，或者把空气压缩进一个容器里）。结果表明一定数量的不论什么形式的功，都产生出相同的热量。这一发现得到了公认，现在能量的最基本单位就是以焦耳的名字命名的[1焦耳大约等于在地球表面附近，把一只苹果垂直举起1码（0.9144米）所需要的功]。

能量的不同形式热和功之间的等价是热力学第一定律的基础。这一定律说，在一个物理过程中，能量总是守恒的，尽管它可以从一种形式转化为另外一种形式。换句话说，如果你对任何一个物理事件拟出一份能量清单，则事件前后的总能量是相等的。唯一的区别是，开始时的能量的一部分或全部，必定会在事件后作为热量出现。这是因为总有某些能量在某种物理过程中被“烧掉”了——例如，克服摩擦和空气阻力。这些“烧掉”的能量不会真的从清单上消失，只是由于表现为废热的能量耗散，使得能量换了一种形式。

在每一个能量转化的过程中，都有因为产生热而出现的能量耗散。例如一位运动员在跑100米赛跑时消耗的化学能，一只白炽灯泡发光（电磁能）时消耗的电能，等等。对耗散问题更进一层的认识，主要是由德国物理学家克劳修斯（Rudolf Clausius）完成的。他生于1822年，父亲是一位牧师兼教员。克劳修斯了解到，虽然热和功在焦耳所表明的意义上是等价的，但是耗散使得它们之间产生了一个十分重要的不对称性。原则上，任何形式的功都可以被完全转化为热。但是耗散意味着相反的说法不能成立，在热转化为功时，总有一部分热白白浪费掉了。例如，并不是蒸汽机中所有的热都可以用来推动活塞。一部分热量浪费到加热机器、周围大气以及操作者的手。一部分热量在机器关掉之后仍然保留在小水滴中。克劳修斯认识到（虽然最

初这一认识还是朦胧的），这意味着热量损失是不可逆的，一旦发生热量损失，这种废能绝不可以再次变成为功。他的这个突破性发现在1850年得到完全证实，克劳修斯也因此被称为“时间之箭”之父。但是在那个阶段他的学说还是有毛病的。出生在北爱尔兰贝尔法斯特的数学家汤姆孙（即后来的开尔文勋爵），把克劳修斯这种笨拙的处理方式改成一种普适的表述，即热力学第二定律。按照这个定律，总是存在着机械功退化成热的无情倾向，而相反的倾向却不存在。

第二定律的含义是，所有的能量转化都是不可逆的。当发动机的曲轴正好转过一个循环时，这台机器回到这样的状态，也许连最能干的机械工也看不出它与原来的初始状态有何不同。但是由于热而产生的能量损耗，却已经使得一些无法再被消除的变化发生了。曲轴规律性的机械运动（功）受到了摩擦力。一部分机械能变成了热，这热我们可以想象为轴上分子无规则的随机运动。这种无规则运动的能量又有一部分会被空气分子带走。最后的结果是作为热而出现的不可逆的能量耗散。

这里我们要记住的一个关键因素是卡诺提出的一个观点，即热只能由较热的地方流向较冷的地方。这就使我们得到了第二定律的另一种表述形式，即不可能通过把热量从较冷的地方转移到较热的地方而做功。为了更幽默地描写一下这个在实际行为中不可避免的原则，我们想引用喜剧大师弗兰德斯（Mi-chael Flanders）和斯旺（Donaid Swann）在他们的歌“第一和第二定律”中的一段歌词：

你不能让热从冷处传到热处，

你想试一试吗？结果只会一无所获。

“不见得吧，让冷的再变冷，

会让热的再变热——这合乎规矩，

就像热量从热的东西传出，

一定会使冷的东西变热。”

不，你不能让热从冷处传到热处，

你如果要试试，你就会像是一个蠢货。

冷的东西变热，这才合乎道理——

因为这是一条物理法则！

克劳修斯在1865年引入“熵”的概念，从而把可逆过程与不可逆过程加以区别。这就使得第二定律更加具有锋芒。熵是这样一个量，它在有耗散的情况下不停地增长，当所有进一步做功的潜力都已耗尽，它就达到了极大值。按照克劳修斯对第二定律的说法，在可逆过程中熵的改变是零，而在不可逆过程中熵总是增加的。熵这个名称是克劳修斯根据两个希腊字发明出来的，意义是“转移的量”或者“发生变化的能力”。它无疑是热力学中最重要的概念，并且给出了一个明显的时间箭头：熵的增加正好与时间的前进一致。熵的概念给了爱丁顿极深刻的印象，他写道：“我希望能把熵这个概念在科学研究中令人惊奇的威力，原原本本地告诉你。”他把熵比作美和旋律，因为这三者都是与排列和有序联系在一起的。

为了更深入地理解熵的意义，把一些杂乱的容易引起混淆的因素去掉是非常有帮助的。科学家们为此常用一种理想化的情况，即把所感兴趣的过程定义为系统，过程以外的世界构成系统的外界。例如，热力学家常喜欢考虑“孤立系统”这样的特殊情况，这种系统与外界完全无关[图13（a）]，像一只有刚性绝缘壁的盒子，物质或者能量无论从哪个方向都不能通过。一只理想的咖啡暖瓶可以作为这样的例子，它不会散失水蒸气或者热能。虽然在实际上没有一个系统是完全孤立的（宇宙本身可能除外），但是这还是一种非常有帮助的理想化

假设。另外还有两种普遍形式的系统在热力学中常被应用："封闭系统"[图13（b）]，它可以与外界交换能量以及"开放系统"，它与外界既可以交换能量也可以交换质量[图13（c）]。利用这样的专门术语，像人这样的有生命的物体就是一种开放系统，因为人要和外界发生能量和物质的交换，例如喝酒、吃肉以及呼出热气和产生排泄物。

按照第二定律，在一个孤立系统中自然发生的任何过程，都一定伴随着系统的熵增加。因而熵给所有孤立系统提供了一个时间箭头。当熵达到它的极大值时，孤立系统的时间演化就停止了，该系统就处于它最无序的状态。这时系统已耗尽了它所有发生变化的能力——它已经达到了热力学平衡。

克劳修斯认识到宇宙本身是一个完全孤立的系统（否则在它外边是什么呢？），并且在1865年把热力学的前两条定律写成宇宙学的形式。第一条定律说，宇宙的总能量是守恒的；第二条定律说，如我们已看到的，宇宙的总熵是在无情地朝着它的极大值增长。德国物理学家霍姆霍兹（Hermann von Helmholtz）第一个从第二定律推断说，整个宇宙的演化就是逐渐地退化，最后停止于热力学平衡，此时不会再有任何变化发生。一个处于平衡的宇宙，熵和无序性都达到最大，所有的生命也就随之而死亡。宇宙的这种"热死"或"热寂"，又引出了弗兰德斯和斯旺的妙句：



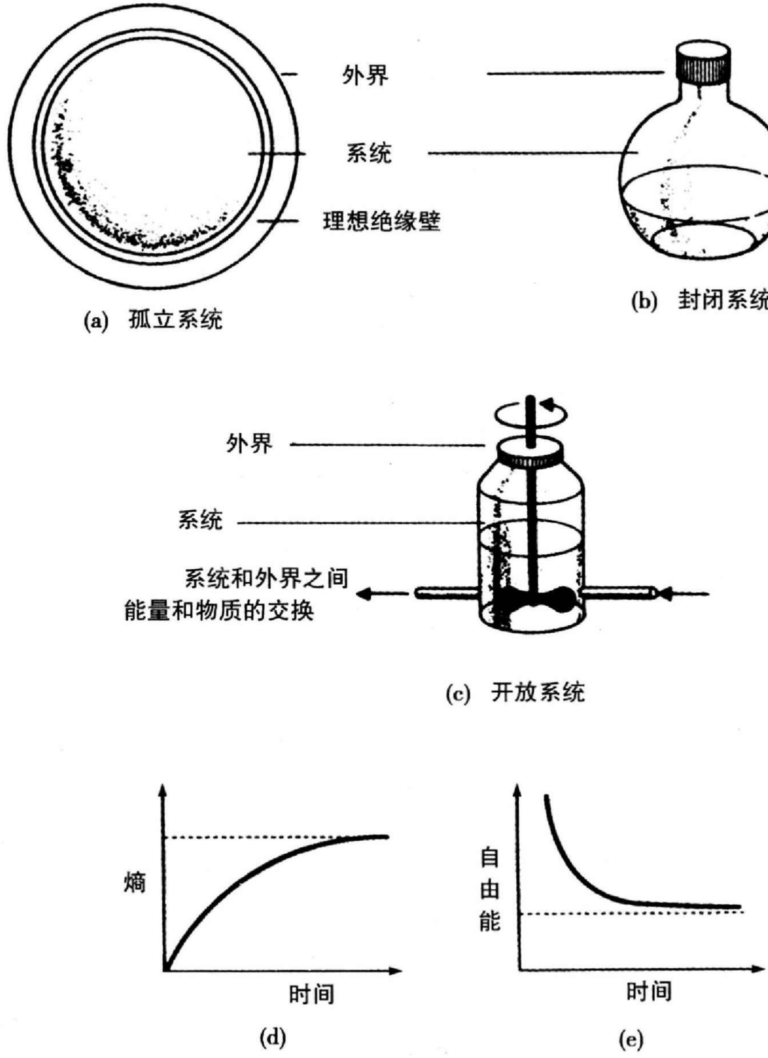



图13　热力学中的三种系统：（a）孤立系统，（b）封闭系统，（c）开放系统，图（c）表示熵的增加，图（a）表示自由能的减少。[录自柯文尼，法文杂志《研究》第20卷，190页（1989）。]

热就是工作（功），而工作是该死的东西。

宇宙中所有的热，

因为不能再增加，

都在逐渐冷下去。

此后，不会再有任何工作了，

将是天下太平、永远的休息。

真的吗？

真的！老兄，这就是熵，

所有这些都是因为

热力学第二定律！

这是一个科学理论从一团蒸汽得到的一幅非常可怕的图像。然而，正如马克·吐温（Mark Twain）所说："先抓住事实，然后你才可以大做文章。"实际上，热寂的说法是有问题的，因为它忽视了引力（和黑洞）的作用：如果包括引力，宇宙必定越来越远地偏离热寂所想象的物质均匀分布。即使不考虑这一点，宇宙这样一幅惨淡的远景也和它的中短期行为没有多大关系。我们从天文事实中（第三、第四章）知道，宇宙作为一个整体是在膨胀着的，所以在任何地方它都不会接近热平衡。热力学的知识也告诉我们，当一个体系远离平衡时，会由于像天空上星星那样的局部热点，而发生一些非常有兴趣的事情，例如生命的出现。

这里我们看到，热力学和达尔文生物进化论之间有一种明显的冲突。达尔文在他的自然选择理论中，表明了大自然何以能够优先选择一些罕见的事件（变种），因而逐渐演化出越来越复杂的生命形式。在他的理论中，变化的推动力是一些随机发生的事件。然而，玻耳兹曼（他是我们这一章的主角）却表明，在充满分子的气体中，正像克劳修斯的第二定律所说的那样，高度有序的结构将随着时间随机地消

失。有些人说，“克劳修斯和达尔文不可能都是对的。”我们将在第六章和第七章再谈到，达尔文的学说如何与热力学取得一致，这里我们只想指出，玻耳兹曼对于达尔文，这个比他年长的同时代人，是十分尊重的。在19世纪将近结束的时候，玻耳兹曼写道：“如果你问我内心深处的信仰，关于是否有一天本世纪会被称为铁的世纪，抑或蒸汽的世纪或是电的世纪，我会毫无疑问地回答说，它将被称为自然机械观的世纪，达尔文的世纪。”1886年在奥地利科学院的一次会议期间，他在关于第二定律的讲演中谈到了同样的观点。

如果科学家们认定达尔文和热力学之间确有无法沟通的分歧，则大多数物理学家一定会说是达尔文错了，因为第二定律已经被证明是普遍适用的。用爱丁顿的话来说，“我认为，熵增原则——即热力学第二定律——是自然界所有定律中至高无上的。如果有人指出你所钟爱的宇宙理论与麦克斯韦方程不符——那么麦克斯韦方程就算倒霉。如果发现它与观测相矛盾——那一定是观测的人把事情搞糟了。但是如果发现你的理论违背了热力学第二定律，我就敢说你没有指望了，你的理论只有丢尽脸、垮台。”

## 平衡态热力学

如果你对蒸汽机感兴趣，第二定律所预言的、耗散引起的热损失就显然是一件讨厌的事。因此，早期热力学家们致力于寻求避免不可逆性的途径。在“思维实验”的理想情况下，这看起来是一件容易的事：可以简单地让蒸汽机无限缓慢地工作，这样在每一个瞬时，系统和外界之间就会处于热平衡。在这种“准静态”情况下，系统的整体性质不会随着时间而改变。熵会在任何时候都处于它的极大值，并且没有不可逆的热量损失发生。这样的系统相应于一个完全可逆的热机。即使如此，如我们已看到的，热转换为功的效率也不可能达到百分之百——第二定律是不可能违反的。同时，这样的理想情况还是有美中不足，因为在上述条件可以实现的情况下，这台机器要用无穷长的时

间才能完成这一最简单的操作。

这些问题也许说明，平衡态的处理是颇有缺陷的。它相当于抑制时间在过程中的基本作用。因为所有的过程都发生在一段有限的时间内，因而在这一过程中不可能包含有无限多个平衡态进程。然而，许多科学家仍然试图按这种有点自相矛盾的方式思考热力学过程。这其中的一个原因是，只讲平衡态过程，可以避免描述不可逆过程的不便和困难，尽管在使一个系统达到平衡之前是要经过不可逆过程的。但事实上，克劳修斯关于熵的定义是仅仅对平衡态而言的，而且，熵是否可以在平衡态之外的一般情况下定义，这个问题现在仍然没有解决。

美国物理学家吉布斯（Josiah Willard Gibbs），无可非议地被认为是现代平衡态热力学的奠基人之一。在把热力学从局限于研究热和功之间的关系，拓展到研究所有形式的能量之间的转化方面，吉布斯做出了主要的贡献。他的3篇关键性的论文发表于1873年到1878年之间，当时他还只有三十多岁。这些论文是由康涅狄格州科学院发表的，后来它的出版委员会承认说，当时他们之中没有一个人懂得吉布斯说的是些什么。有一个人这样说："我们认识吉布斯，信任他投来的稿件。"吉布斯和他的许多后继者，都谨慎地避免提及非平衡态现象。虽然他们的研究对于已经达到平衡态的系统有至关重要的意义，然而还是有点像把医学限制为给死人看病。

大家都说，吉布斯是一个为人死板、极不活跃的人。即使是一些高级学者，也常常把他和当时美国杰出的化学家沃尔考特·吉布斯（Wolcott Gibbs）相混淆。很长时间以后他的学说才在学术界得到广泛认可。今天，他的学说使我们感兴趣的主要是，能帮助我们洞察究竟是什么使得系统停止随时间的演化。当一个孤立系统[图13（a）]，例如盛在一个充分绝热容器中的气体，处在它最无序的状态时，它就已经达到了热力学平衡。此时，为了描述平衡的宏观状态以及演化的终点，所需要的只有唯一的一个量，即熵可能达到的最大

值。但是对于封闭系统和开放系统[图13（b）和图13（c）]来说，它们与外界的交流越来越重要，因而熵最大的状态也必须同时把外界的熵计算在内。这样，对于厨房里一杯正在凉下来的咖啡，我们必须考虑咖啡和厨房之间的能量交换对厨房产生的微小加热。如果我们想单独研究咖啡的平衡性质，这就麻烦了。因而为了简单起见，我们总是希望避免把厨房的行为明显地带到讨论中来。

为了排除厨房的影响，我们可以引入一个叫做自由能的新的量，它在平衡态时具有最小值。一个系统的自由能代表能从该系统得到的最大有用功。虽然自由能只是总熵换了一种形式，但是它的好处在于，它可以被看做是咖啡的某种内禀性质，而无须涉及厨房里其他地方发生了什么情况。在整个物理学和化学对系统平衡性质的描述中，自由能都起着中心作用，不管该系统是磁性物质、电冰箱还是化学反应的混合物。

熵和自由能是所谓热力学势的例子。这指的是，它们各自的极值——熵的极大值和自由能的极小值——显示着热力学平衡的状态。可以用钟摆的摆动作为一个比拟。在每一次摆动开始的时候，摆锤在引力的作用下具有一个势能，当它沿着弧线往下摆动时，势能就转化为运动的动能。空气的阻力使每次摆动开始时的势能逐渐消耗掉，因而摆动的弧线越来越小，最终使摆锤停止在垂直的位置上。这就是它势能量小时的位置。在一个密闭的容器中发生的化学反应，情况也与此十分相似。它最终会达到自由能最小的状态，此后不会再发生进一步的变化。

一个更时髦的描述热力学势极值的方式，是说它们是系统在时间上演化的“吸引子”。当一个球向山谷滚下去的时候，无论滚动开始的位置如何，球最后停下来的地方总是不变（图14）。与此相同，一个化学反应的平衡态决定于熵最大或自由能最小的点，与反应开始的条件无关。

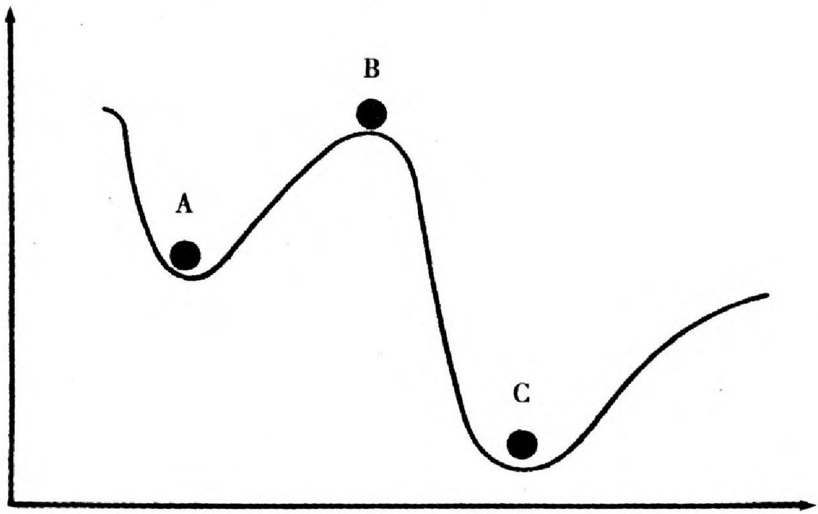



图14　热力学（熵和自由能）的力学类比。一个力学系统在势能最小时达到平衡。位置A和位置C相应于局部势能极小值。球B点是不稳定的——哪怕是最小的扰动也会使它滑到A或者C，这个过程中就会有能量的耗散。

平衡吸引子标记着一个重要的停止点，就像一块路标那样，“一切变化终止于此！”也像死亡对我们大家一视同仁，不管生前富贵荣华还是穷困潦倒，死亡都会把每一个人同样带到坟墓，这些吸引子也把它们的系统无情地拖向平衡。吸引子有各种各样，我们以后会常遇到。一个化学反应的平衡吸引子可以比作一个漏斗形物的底部：无论一个球开始时的位置如何，球将总是滚到最底部的一点。但是并不是所有的系统都有像这样的一个简单的吸引子——打个比方说，其他的（非平衡态）吸引子可以使这个球沿着一个一维的环滚动，像沿着宽边帽的帽檐滚动一样；或者使这个球在更高维的空间中更大的范围内滚动，像沿着一个油炸面圈的环形表面滚动一样。我们以后会看到，还可能有更稀奇古怪的吸引子。

## 砸碎平衡

平衡态热力学对于研究宏观系统不随时间而变的性质非常有用

——例如像对一个所有化学反应已经停止了的试管，灰烬，或一杯已经凉了的茶。吸引子的概念同样可以给我们提供一个“时间之箭”的“箭靶”。然而，在非常实际的意义上，平衡态同样是一条死胡同。因为它涉及的只是热力学演化的终态，因而也就是时间的终点，它不可以用来描述能使时间明显表现的过程。如兰兹堡（Peter Landsberg）所说：“……热力学中既然不出现时间坐标，因此就不会有任何事情发生。”真实的世界中很少有这种令人窒息的平衡态。文学家贝凯特（Samuel Beckett）写道：“我们在呼吸，我们在变老！头发脱落，齿牙动摇！青春一去不回，往事云散烟消！”对于有生命的物体，热力学平衡只是随着死亡而出现，那时腐烂的尸体最终化成腐土。生命是由许多过程所组成的，从细胞分裂、心脏跳动到消化和思维，所有这些过程的发生都是因为它们不是平衡态。一个更有意思的说法是，平衡态热力学所做的预言，就像一个算命先生对你说你将来会死，但是他根本不知道你马上要和你的情人约会。对于一个基于第二定律及其对变化的解释的理论，这样一个结论看来是非常令人失望的。为了使我们对时间的研究有进一步的进展，我们必须再调查，对于不是平衡态的不可逆过程，热力学可以告诉我们些什么。

非平衡态热力学自然分成两个分支：“线性”分支描述接近平衡的系统的行为，“非线性”分支处理系统远离平衡时的情况。“线性”是一个数学词汇，它说的是在任何一种现象中，只要两个量之间有正比关系，这种关系就可以在图上用一条直线来表示。一个线性系统的作用，是组成该系统的各个部分的作用之迭加；但是一个非线性系统的作用，就不是简单地把各个部分的作用加在一起了。这之间的差别，可以用通常的家用电炉和切尔诺贝利核电站四号反应堆来比较，这后者是迄今世界上最严重的核事故的源地。如果某人有2台电炉，则他会得到双倍的热量；如果他有3台，则热量会是3倍；依此类推。这种情形表示，在产生的热量和电炉数目之间有一种线性关系。但是在切尔诺贝核电站的情况下，反应堆芯的过热增加了链式核反应，这反过来又产生了更多的热。结果形成的汹涌澎湃的热流，完全与最初的温

度升高失去比例，这就叫做非线性正反馈。草率的操作规程加之（前）苏联式设计中的致命弱点，最终导致了能量爆炸性的释放。1986年4月26日核电站变成废墟，并产生了全球性的严重后果。

## 混沌中现出有序

宏观系统的自然倾向是循着时间箭头走向平衡态。但是，如果系统的这一过程在它达到目的地之前就停止了，情况又会是怎么样呢？

图15所示的盛有氢气和硫化氢气体混合物的容器，可以用来说明，只要使一个系统保持在非平衡的状态，就可以推翻“熵就是无序”这样一个肤浅的教条。本来我们会以为，当混合物被加热时，它会变得更加无序：热量加得越多，气体分子就会越起劲地在容器里到处乱跑。但是现在让我们来防止它达到平衡态，这只要使容器的两端保持一个很小的温度差就行了。实验显示，在容器中的两种气体将逐渐发生分离：容器中较热的一端富于较轻的氢气，而较冷的一端高度集中着质量较大的硫化氢。这个效应称为热扩散，这同一效应也使得储油罐中的油物质出现类似的浓度分布轮廓。

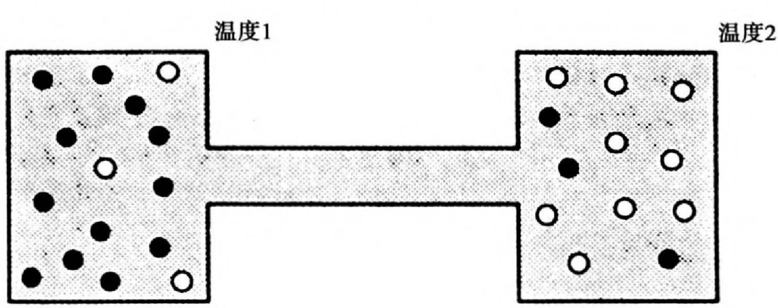



图15　热扩散装置。如果两个容器之间保持温度差，在其中一个容器中会发现更多的某种分子，而另一种分子则较少。浓度梯度正比于温度差（$T_1$-$T_2$）。[根据普里高津等所著《在时间与无穷无尽之间》第50页。]

初看上去这个现象是与直觉相反的——加热容器使得熵增加，然

而却使分子分布的随机性减小了，显现出一个所谓的气体浓度梯度。尽管热力学第二定律的通俗解释，是熵直接联系“无序”，而热扩散却表明，有序的组织可以自发地从无序状态中形成。在上面这个简单的例子中，这种有序组织就是沿着温度梯度方向，氢气浓度的逐渐增大，硫化氢浓度的随之减小。诚然，在气体分子的激烈运动中还是存在有随机性，但是总的说来这显然不是处于热平衡。热扩散给出了第一个例证，说明在不可逆的、非平衡态过程中，可以产生出有序性。这样，时间箭头就和可能出现的结构联系了起来。

平衡势使我们能够预测一个热力学系统的最终状态。它表明系统就像一个在路上滚动的球，无论它从凸起的什么地方开始滚动，最后总是停止在低洼之处。当一个系统由于其周围环境的限制而不能达到平衡时，是否对非平衡态行为也有一个类似的说法？总的说来答案是肯定的，只要系统不是偏离热力学平衡太远：还是用球来作比喻，球总是会滚到低洼处的。但是它也许会滚到另外什么地方——例如，如果它跑得离低洼处太远的话，它会落到另一个坑穴里面，这就使得预测变得困难得多。为了对热扩散确定一个热力学势，容器两端的温度差，也就是温度梯度，必须充分小。

温度梯度给了该系统一个“推动”，因而可以被描述为如同一种热力学力。这样的力造成了热量流和质量流，就像踢球使球运动一样。在接近平衡的时候，热流和质量流表现得很简单——如果力增加一倍，则流也增加一倍，如此等等。热流和质量流与造成它们的力直接成正比，因而我们把这种情况下的热力学叫做线性热力学。

线性热力学主要是由耶鲁大学的昂萨格（Lars Onsager）的努力，才在20世纪30年代打下了坚实的基础，昂萨格为此获得了1963年的诺贝尔化学奖。他的“倒易关系”表明，在线性系统中有一种美学上很漂亮的对称关系：力产生流，流也产生力。在热扩散的情况下，物质的流动是由一个力引起的（用热梯度来表示）。昂萨格倒易关系接着说，物质的浓度梯度将产生热流，这个效应已经在实验上得到了

证实。

1945年，一个新的角色登上热力学舞台——这就是普里高津（Ilya Prigogine）。他在1917年即俄国十月革命那年出生于莫斯科，十年以后，小普里高津随同家人迁往西欧，在布鲁塞尔定居下来。在由比利时物理化学家顿德尔（Thophile de Donder）创立的布鲁塞尔热力学派里，他以一个小学徒的身份开始，一步步成长起来。他的大部分大学生涯是在布鲁塞尔自由大学度过的，这所学府当初是由互济会建立的，为的是对抗天主教会对纯世俗事务的横加干涉。至今这仍然是这所大学的宗旨。

28岁的普里高津发现，在线性表现良好的区域，热力学耗散降到它可能的最低点。这样，系统熵的变化率，也就是内禀的熵产生将会减小：一般地，系统将会演化到一个稳定的或不变的状态，此时耗散处于一个极小值。在热扩散的情况下，总的熵可能是增加的，但是当气体最终的浓度梯度已经建立以后，内禀熵的产生率就处在它的最低值。

普里高津的博士论文在1947年发表，题目是“不可逆现象的热力学研究”。它包括有最小熵产生定理，并且也为他今后一生的研究打下了基础。普里高津像一颗新星那样升起，最后成为布鲁塞尔学派的领导人物。他和他的老同事格兰斯多夫（Paul Glansdorff）一起探索把热力学分析扩展到新的领域。

普里高津的最小耗散的图像，比起最大熵的平衡态概念来对我们更为有用，因为它与实际世界的关系更为密切——在实际世界中，没有东西是真正处于平衡态的。进一步演化的趋向总是有的：液体会混合，建筑物会风化，物体会冷却。但只要有一个很小的外部影响，使得系统保持在偏离热平衡的状态（例如在发生扩散的容器中的温度梯度），则将持续出现的是一种“稳恒态”，而不是坍缩到完全无序的状态。大多数房主是熟悉这种稳恒态的。风吹雨打和风化作用早晚会使

得房屋变成一堆瓦砾，这就是它的平衡态。但是房屋通常可以在许多年里保持在稳恒态，因为维修的速率等于损坏的速率。只有当房屋停止维修时，房屋才开始风化瓦解（虽然最后的坍塌是很多年以后的事，但无论如何早晚一定会发生）。

在平衡态热力学的情况下，时间之箭的“箭靶”可以用一个具有像最小自由能或最大熵这样一些量的固定点吸引子来描述，它把系统拉向平衡态。对于偏离平衡，但是由于外部影响相对很小，而使得系统能够保持接近平衡态的情况，我们可以看到，许多平衡态情况下发生的事情仍然成立。最终的状态——例如像浓度梯度——仍然不随时间而变，仍然是恒定的。普里高津把分析推广到略微偏离平衡的情况，他发现系统通常会演化到熵产生为极小值的点。此时他希望大胆地跨进一步，把这样的分析用于更复杂的情况——特别是远离平衡态的非线性系统——以完成一幅更大范围的系统随时间演化的图像。然而，与平衡态行为的类似性此时不复存在。这个新的领域也许很难懂，但是却非常令人兴奋，因为它和我们所看到的周围世界有非常密切的联系，并且提供了关于时间和变化的一幅更深奥微妙的图景。

## 极度远离平衡态

普里高津的最小熵产生定理是一个重要的结果。然而，它的证明决定于昂萨格所描述的流和力之间的良好线性关系（以及别的一些因素）。普里高津和他的布鲁塞尔同事一起，决心探究远离平衡态、线性规律已遭破坏的系统，目的是发现是否可能把他的定理推广为一个演化判据，对流和力之间的简单关系不再成立的非线性系统也能适用。这一努力导致了一些非常具有魅力的结果，也引起了不少激烈的争论。

我们前面已经提到过，科学家像其他人一样，他们之间会发生激烈的争吵甚至敌对。围绕着布鲁塞尔学派的争论一直没有平息，不幸

的是，这个小组向外界报告他们的发现时，偶尔也做过一些夸张，这使得争论如同火上浇油。其实根本的问题主要是介绍的方式，但却导致了非常令人不快的结局：差不多像政界的观点两极分化。可悲的是，这种宗派之争逐渐破坏了对布鲁塞尔学派做的贡献的理性辩论的基础。

经过长达20多年的煞费苦心，布鲁塞尔学派精心研究出一种称作“广义热力学”的理论，它的目的是把热力学原理运用于远离平衡态的情况。本质上，格兰斯多夫和普里高津用的是一种近似，它使得一个远离平衡态的系统，在局部上表现为平衡态。整个系统好像由许许多多这样的局部拼缀而成。从概念上讲，它颇为类似广义相对论中，把弯曲时空想象为许多局部平直时空拼缀在一起。

格兰斯多夫和普里高津用这种“局部平衡”近似，来研究平衡态热力学远不能够处理的情况。他们系统地阐述了所谓的“普适演化判据”，而且认为这一判据会对事物随时间演化的方式，给出一个更加精深奥妙的见解。这一研究在他们于1971年出版的一本书里登峰造极，这本书名叫《结构的热力学理论，稳定性和涨落》。不幸的是，他们用以表达这个判据的措辞——特别是他们选择的“普适”这个词——又成为后来争论的中心。作者们的原意是，他们得到的是一个纯热力学的演化判据，但是由于他们所声称的看上去比实际成立的要多，因而使得他们自己成为一场辱骂性论战攻击的目标，对手是那些仅仅是急于对布鲁塞尔学派发难的人。

格兰斯多夫-普里高津判据处理的是我们可能遇到的最普遍的情形——远离平衡的开放系统，其中能量和质量都可以流动，系统的行为由非常复杂的非线性关系所支配。对于远离平衡的稳恒态的稳定性，他们的判据给出了一个总的说明。这判据说，当平衡偏离得太远时，这种稳恒状态就变得不稳定了。此时可能会出现一个“转折点”，或者称作“分叉点”，系统在此处会偏离稳恒态，而演化到某种其他状态（图16）。

一个重要的新的可能性是，在第一个转折点之后，系统在时间和空间上的行为会具有高度的组织性。例如，在某些远离平衡的化学反应中，我们可以看到出现规则的颜色变化或者漂亮的彩色涡漩。这是系统随时间演化过程中达到的稳定状态，但是已经不再和最小内禀熵产生有任何关系了。

为了更清楚地说明熵和有序行为之间的关系，也许我们可以用一位打算省钱的先生（当然一位女士会做得同样好）可取的选择作比方。如果他是一文不名，那么他在消费上的改变——相当于平衡态时熵的变化——是零。然而，真实的世界不是处于平衡态。所以，我们这位希望省钱的先生，非常可能至少有少量的钱，而且必须花费一些在饮食上。在这种情况下，他能够期望做到的是，把他的开销降低到能使他活下来的最低水平。与此相似，许多处在少许偏离平衡但还是稳定状态的热力学系统，将把熵产生降到最低点。现在让我们来看远离平衡时会发生什么情况。这相当于这位先生是一个已经有钱而想要更多钱的人。在这种情况下，把他的日常开销降到最低是远远不够了。现在对他来说有许多其他的选择，而不仅仅是选择最便宜的食品和住宿。他可以把钱投资到利息高的银行里。他也可以斥巨资去进行利润丰厚的投机冒险事业。所有这些短期花费都完全可以使他达到他的目的。此时对省钱来说（而不是对花钱来说）不再有一个普适的判据，而如果他是一个过着粗茶淡饭日子的穷汉，就会有这样一个判据。这就是说，对远离平衡的情况，呈现出许许多多有序行为的可能性。

值得注意的事实是，当系统远离平衡时，整体熵产生以极快的速率增长，这是与第二定律一致的，然而我们却同时观察到极其有序的行为的出现。这样，我们必须修改前人留给我们的信条，即时间箭头总是联系着向无序状态的退化。确实，当时间走到“终点”，再不会发生任何变化的时候，可能出现的就是无序状态。但是在较短的时间尺度上，我们可以看到短暂的结构的出现，只要物质和能量的流动可以

继续，这种结构就可以维持存在。下面这一点确很重要，这就是只有当系统对外界是开放的，它才可以保持在偏离平衡的状态：因为这才使得系统所产生的熵可以输送到外界，从而系统可以维持在有序状态，同时容许系统和外界所构成的整体熵增加。

格兰斯多夫和普里高津所给出的是一个“弱”判据，因为它只是说了热力学转折点的可能性，而没有给出它的必然性。它没有像热力学第二定律那样，对宇宙万物随时间的演化方式给出一个铁的法则。它依赖于第二定律以及其他一些附加假设。远离平衡的时候一般地不再有热力学势——不再有作为时间之箭箭靶的单一吸引子。这个判据缺乏普遍性这一点，是美国加州大学戴维斯分校的凯瑟（Joel Keizer）和亚特兰大乔治技术研究所的福克斯（Ronald Fox）指出的。

## 生机在远离平衡态时萌动

像《银光先生历险记》中那只不叫的狗，给福尔摩斯提供了破案的极其重要的证据一样，格兰斯多夫和普里高津判据的短处，可以转化为值得重视的长处。这个判据之所以不能成为一个普适法则的原因，是由于对远离平衡态的情况，有多种多样的可能性可供选择。这种彻底的复杂性，使得直接因果联系的描述变为不可能，同时可以帮助理解，为什么在许多可能性中采取一种，就赋予系统在时间上演化的实际途径一个特殊的地位。这可以借助于所谓分叉图来说明，它是一种简单的形象化表示，描述了可供选择的多种多样的可能性（分叉的意思是在转折点有叉状分支）。对于我们这位节俭的先生所处的情况，图上表明在接近平衡的时候只有一种可能性，这相当于他处在捉襟见肘的贫困线上（相应于热力学主支）。远离平衡的时候，会出现多得多的可能性或“分支”。图16（a）中，分叉表示当系统达到它的热力学转折点时，该系统面临的难题。

让我们把这种处理用于化学反应，这是不可逆过程的一个极好的

例子。随着时间的演化，化学物质相结合，并且从而转变成化学性质上截然不同的产物。例如，氧和氢结合在一起生成水，而铁和大气中的氧相作用形成铁锈（铁的氧化物）。分叉图上显示了，当一个特别的化学反应远远偏离平衡时，可能出现的情况。图16中的纵轴代表反应混合物中某种化学成分A的浓度。横轴代表与化学反应停止时热力学平衡态的“距离”（用希腊字母λ来表示）。点A★（A的平衡态浓度，此时λ等于零）是系统中A在平衡时的浓度。当一个系统由于某种限制而不能达到平衡时（例如一个开放系统，其中生成物的排出可以不断地由外界物质的输入加以补充），A的浓度将与A★不同。在这种情况下，反应剂A的补充速度与被反应消耗的速度相同，因而使得A的浓度保持不变。这时的浓度与A★的差别，就可以作为系统保持偏离平衡时，与热力学平衡态的“距离”的量度。

分叉图给了我们一些什么启示呢？在平衡态以及平衡态附近，λ处于线性良好的区域（即在热力学主支上），并且系统遵循普里高津的最小熵产生定理，反应混合物也处于稳定的稳恒态。但是当与平衡态的距离超过一个临界的阈值时，就会出现选择，也就是在热力学转折点，即分叉点$\lambda_c$，突然出现两条线。图上的虚线表示热力学主支，现在它变成为不稳定的了。

这种转折点具有特殊意义，因为越过它以后（它的位置决定于所研究的化学反应的细节），我们在有些情况下可以看到有序行为的出现。这个区域的另一个特点是，此时化学反应可以具有选择。在这种特例中，化学反应有两个新的稳定状态可供选择，过了线性区域后，A的浓度在这两种状态下的值是根本不同的。转折点实际上把一种演化历史感引入到反应混合物中来。这是因为，在许多选择中间必须做出一个选择，使得系统能够沿着分叉图上的一定路线演化到目前状态，因而事实上，系统对于它所做出的选择是有“记忆”的。在另外一个容器里生成的混合物可能是不同选择的结果。化学反应在任一时刻只能存在于一种稳定的状态：两条新分支中实际沿循了哪一支，即系

统会处在哪一个新状态，完全像抛硬币决定正反面一样，各有百分之五十的机会。我们将在第六章和第七章看到，当系统面临如此众多的选择，也就是相毗邻的稳定态的数目非常巨大的时候，会出现什么样的情形——这时会发生完全不可预测的动力学行为，也就是所谓的决定性混沌。这种行为可以从图16（b）所示的更广延的分叉图上看出，在这个图上，第一个转折点后面跟着一大批新的转折点。

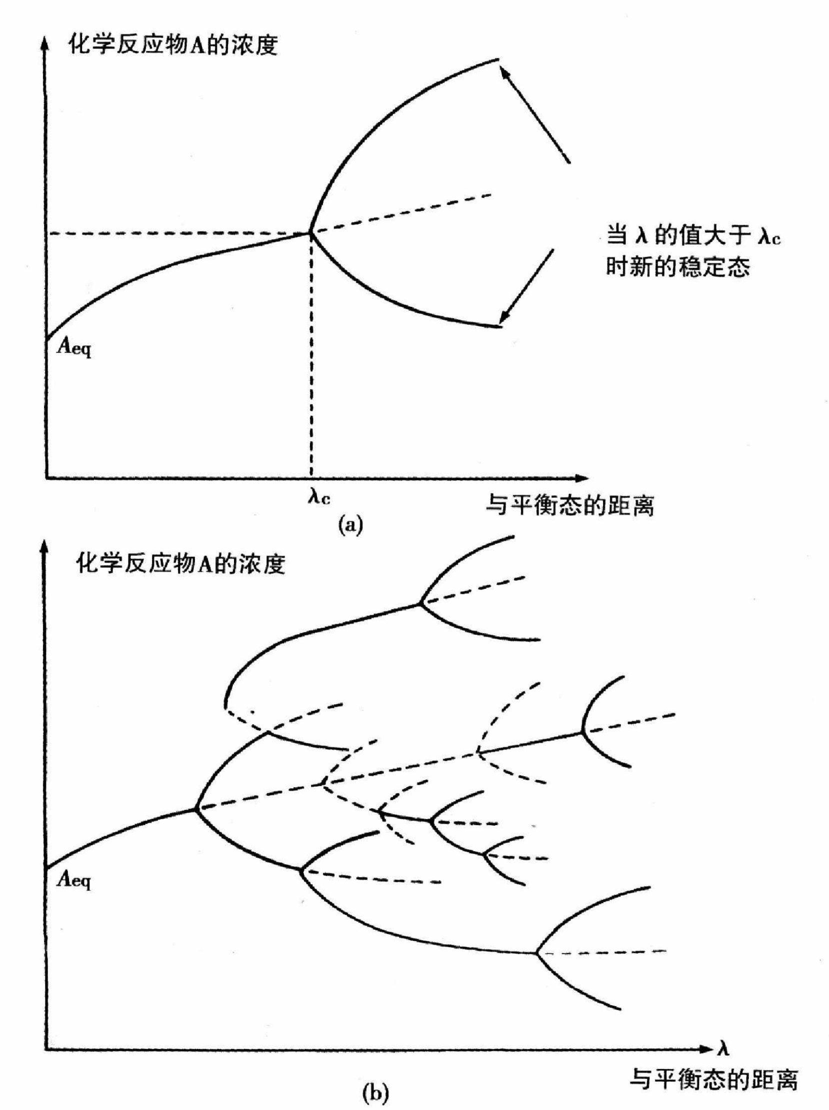



图16　远离平衡时化学反应的分叉图。（a）最初的分叉。λ表示当初最小熵产生的热力学分支变为不稳定时，到平衡态的距离。分支点或转折点决定了λ处的浓度。（b）整个分叉图。当非线性反应远离平衡态时，可能的稳定态的数目急剧增加。

对于在第一个转折点之后会发生什么情况，格兰斯多夫和普里高津的非平衡态热力学只给出了一幅模糊的图像。它暗示某种重要的情况会发生，但对于实际会发生什么，它一点也没有讲。对于一个化学

家，这些选择可能意味着一只周期性改变颜色的化学钟，或者试管里显现出的彩色图案。对于一个生态学家，这些可能性也许意味着动物种群中新的稳恒态或交替变化。对于一个医生，它也许意味着一次心脏病发作。为了恰当地处理这些情况，我们必须基于对（例如描述某个化学反应的）不可逆动力学方程的详尽数学研究。这样的动力学方程引人注目的性质以及它们所描述的复杂纷纭、千变万化的作用，是我们将在第六章要讨论的主题。

远离平衡时出现的新的状态，可以具有一种令人惊异的有序程度，此时无数个分子在时间和空间中的行为达到协调一致。普里高津把这称作“耗散结构”。因为它们发生在系统和外界之间有物质交换和能量交换的情况下，同时伴有系统的熵产生（耗散）。这些导致耗散结构生成的复杂而相互依赖的过程，共同的名称叫做“自组织”。

现在我们可以看到，热力学并不禁止有序结构的自发产生，而第二定律却被普遍错误地认为是朝着无序状态单调地退化的一个代名词。这就是格兰斯多夫和普里高津研究结果的真正意义。1977年，普里高津获得了诺贝尔化学奖，颁奖书中赞扬了他对于非平衡态热力学的贡献，特别提到了耗散结构理论。

如何使耗散结构与“热寂”的概念相符合呢？如果我们的宇宙观草率地应用了平衡态热力学，并忽视了引力的作用，它就会意味着宇宙的演化同义于无情的无序性增长，最终一切变化停止而达到热平衡。但是对于非平衡过程的研究已经表明，一个远离平衡态演化的宇宙，是不能够用如此简单的方式来描述的。在这样的宇宙中，不可逆的非平衡态热力学允许产生自发的自组织结构，使得行星、星系直到细胞、生物得以出现。

## 不可逆性佯谬

许多科学家认为，他们只能理解那些显示出与原子和分子活动有关的现象。这叫做“归并化”，按照这种说法，其他一切都可以归结到这个被认为是更基本的层次。但是用这种方法处理热力学和它的深奥微妙的时间箭头时，问题就出现了。因为如我们已经看到过的，不论是用牛顿力学来描述，还是用相对论或是量子力学来描述，在微观世界里看来不存在时间箭头。这样的一个结果就是，不存在微观平衡状态：时间对称的规律，不会挑选出一个一切变化潜力都已耗尽的终态，因为所有的时间瞬时都是彼此等价的。原子和分子领域中，运动的明显时间可逆性，是直接与热力学过程的不可逆性相抵触的，这就是罗史密茨（Josef Loschmidt）与其他一些人于1876年提出的一个著名的不可逆性佯谬。它认为，热力学和力学这两座大厦充其量也是不完备的。这个问题明显地需要答案，不应置之高阁。我们的科学探索，也许有一天会使我们理解瞬息万变的意义，但是我们不能就此止步，而必须继续努力，直到我们的理解与前面几章所讨论的理论相一致——因为正像开尔文勋爵有一次指出的，一个无视时间箭头的微观世界，具有可以把宏观世界整个颠倒过来的力量。“如果宇宙中的每一个物质粒子，其运动在任何时刻都是完全可以逆转的，则自然界的进程将永远是简单地可逆的。瀑布脚下飞溅的泡沫会重新聚在一起而流回水中；热运动会把瀑布水滴的能量重新会聚，使水升高从而在瀑布旁边形成另一个向上喷的水柱……活着的动物越来越年轻，它们只知道将来而不记得过去，最后会回到未出生的时刻。”

## 跨越鸿沟

要跨越现实世界的这两个层次（宏观和微观）之间的鸿沟，就要用到“统计力学”这门学科，它的任务是把量子力学或经典力学中所需要的庞大数目的信息，缩减到少数几个热力学参量。例如，它告诉我们，如何把对1升气体中的单个分子的描述，代之以气体的压力和温度这类更一般的术语。

在微观层次上，这一升气体由巨大数目运动着的分子组成——这个数目的数量级是10后面接着23个零。因此可以想象，要把所有这些分子的行为总括起来得到一个完整的描述，需要多大的信息量。我们需要知道在某个给定的初始时刻，所有分子的速度和位置。但是在宏观层次上，我们知道只需要很少的信息就可以描述气体的总体性质。例如，在平衡态的情况下，只需要知道三个量——气体的压力，它占据的体积以及它的温度。显然，这使得信息量大大地缩减。统计力学的任务，就是表明如何才能实现这样的缩减。

这其中的关键步骤，是把概率论补充进力学定律中。概率论使我们可以把计算建立在平均值的基础上。这种处理方法的根据有二。其一是，即使对牛顿模型也必须用几率描述，因为我们无法掌握所有分子速度和位置的准确信息（对这一不确定性的深入讨论见第八章）。这样，我们所有的论述就变成统计性的，就像在量子力学中那样，虽然这其中的理由与量子力学完全不同。而且，现在我们也只能谈及系统的平均性质，例如，分子的平均能量。

第二个根据是，宏观系统比起构成它们的原子和分子来要大得多。假如我们想要验证一下，抛硬币时正面（或反面）出现的机会是不是50%。这个几率并不意味着，如果把硬币抛两次，必须出现一正一反。只有当我们同时抛掷大量的同样硬币，或者把同一枚硬币反复抛掷多次，平均的行为才会出现——抛掷的次数越多或者硬币的数目越多，才越接近于50%的几率。回到用平均值来描述一个日常物体的分子的行为来，我们算是很幸运，因为这样的物体所包含的分子数目，可以达到亿亿亿个。因而统计力学中计算出来的平均值，一般能够给出一个极好的描述——与此相同，如果我们把一枚硬币抛掷一亿亿次，那么平均值与百分之五十的几率的偏离就是微不足道的了。

统计力学的一大问题是，它对于一切变化都停止了的平衡态现象的描述很简单，而对于仍然在随时间演化的非平衡态现象来说，就要远远复杂得多。我们先从平衡态统计力学开始，它给出分子与热力学

性质之间的一个明确关系，例如气体分子的平均速度和气体温度之间的关系。平衡态统计力学的关键是一种叫做“配分函数”的复杂数学工具，它能够计算一个系统处于平衡态时的所有宏观性质——例如它的熵和压力，不论这系统是固体、液体还是气体。配分函数在原则上可以准确地从该系统可占的能级计算出来。这些能量可以通过解第四章中谈到过的薛定谔方程而得到。但是从实际的观点来看，这又是一件极其棘手而令人头疼的工作，只有借助于某些奇招，例如巧妙的数学近似，或者是构造一个理想化的模型，才能克服这些困难。然而，一旦我们得到了配分函数，我们就可以用它来给出有关物质所有平衡态性质的答案。例如，可以计算1升气体在平衡态时的熵或自由能，并且确定它在给定温度下的压力，而不需要做任何实验。

当我们用这同样的方法来处理非平衡态过程时，例如描述当容器打开、气体逸出时的行为，就会遇到很大的麻烦。这时候配分函数一下子变得毫不相干，因为只有对于热平衡态它才具有严格的意义。这种简洁巧妙的处理熵和其他热力学量的办法，原来是一条死胡同。

不幸的是，基本原理似乎告诉我们，除此之外，没有其他捷径。让我们来仔细研究一下统计力学的基础之一——即采用平均值，而不是采用气体中每一个分子的准确位置和速度。取平均值的方式，决定于我们选用经典力学还是量子力学来描述微观层次。事实上我们是没有选择余地的，因为只有量子力学才能给出对微观世界的最好描述。然而，尽管这两者之间存在差别，但它们所给出的宏观描述，其基本要点却很少有实质性的不同，除掉某些不寻常的现象，例如超导，即在低温下金属完全失去对电流的阻抗。（这种相似性的出现，是由于经典统计力学和量子统计力学在数学形式上非常相似。实质上，由于所考虑的系统中粒子数目是如此巨大，这就使得行为怪异的量子效应变得不明显了。）

求解牛顿方程式，需要先确定一个大的系统中每一个分子的位置和速度。为了避免这样一个毫无希望的尝试，我们采用一种叫做“几

率分布函数”的统计学方法。像在选举日那天对投完票的选民进行的民意调查，可以告诉我们选举的结果将会如何一样，几率分布函数亦可以告诉我们系统处于一个特殊状态的几率，在这种状态下所有的分子严格地具有所给定的位置和速度，而这位置和速度是按照牛顿方程随着时间演化的。

如果我们更喜欢用量子力学，则可以用波函数，它“等价”于经典力学中的轨道。当然，像牛顿模型一样，量子描述同样有在原子和分子层次上信息量过大的问题。对于一个系统观测到的宏观表现，可以有难以计数的微观状态与之相容（每一种状态由一个单独的波函数来描述），所以我们不可能希望得知系统实际上是处在哪个状态。对于通常含有万亿个分子的气体，量子态的数目之多是无法想象的。因此我们必须转而采用类似于民意调查，或经典几率分布函数那样的量子力学方法。这种方法叫做（几率）密度矩阵，它告诉我们发现系统处在任何一个量子态的几率，而不同的量子态由不同的波函数来描写。

至此为止，看来一切顺利。我们已经设法找到了一条途径，用量子力学或是经典力学来描写大量分子的统计行为。现在我们必须要做的，是描述一个系统偏离平衡态时的不可逆的演化。我们所需要的是描写经典的几率分布函数和量子密度矩阵如何随时间演化的方程。正好这两者的演化方程是一样的，它们称作刘维-纽曼方程。可惜，刘维-纽曼方程背离了我们的目的。它们是直接建立在经典力学和量子力学的基础之上的，而我们知道，这两者都并不区分时间的方向。因此，这些方程同样是时间对称的：它们可以用来计算平衡态时的熵，就像利用配分函数所做的那样；但是它们不可以独自解释系统偏离平衡而演化时熵的增加，而正是这种熵的增加给我们提供了时间箭头。这就是非平衡态统计力学的基本问题之所在。

## 玻耳兹曼的时间箭头

我们在序言中就已经谈到过玻耳兹曼，对这位大师是值得特别地大书一笔的，因为他曾试图解决前面提到的令人困惑的佯谬。他是维也纳一个税务官的儿子，出生于1844年2月20日，即4月斋开始之前的星期二的夜里，那时狂欢舞会已近结束。他后来常常开玩笑说，这就是为什么他的性格如此容易激动、悲喜无常的原因。玻耳兹曼第一个对基本的物理定律——热力学第二定律——给出了一个统计性的解释，但是他的想法遭到其他物理学家和数学家的激烈反对。在他的学说被广泛承认之前，他自杀了。然而他也有强有力的同盟者。1900年，即玻耳兹曼自杀前仅仅6年，21岁的爱因斯坦写给他的女友米勒娃说，玻耳兹曼“是一位高明的大师。我坚信他的理论原则是对的，也就是说，在气体的情况下，我们确实是在处理具有确定的有限大小的分立质点，这些质点遵照一定的条件在运动”。

作为一个虔诚的天主教徒，玻耳兹曼在奥地利的萨尔兹堡和林兹接受教育，那时正是奥匈帝国时代。当他在维也纳大学攻读数学物理学博士学位时，他就显露出罕见的数学天才。他的导师是物理学家斯忒藩（Joseph Stefan）以及脾气古怪的数学家佩兹瓦尔（Josef Petzval），他是斯忒藩的博士生导师，是他把麦克斯韦的气体动力学理论引进欧洲大陆。玻耳兹曼把麦克斯韦的工作做了进一步的推广，他用的是斯忒藩的另一个学生、前面提到过的罗史密茨最早提出来的统计力学方法。早在量子力学出现之前许多年即1872年，玻耳兹曼就报告了把刘维方程用于气体中大量分子集合的结果，并分析了处理热力学平衡的途径。他的研究终于使时间似乎在微观层次上有了方向。他得到了一个时间不对称的演化方程，现在叫做玻耳兹曼方程，此方程对所谓的单粒子分布函数成立，这种分布函数是对气体中单个分子运动的一种统计描述。从这一方程他构造出一个新的数学函数，即所谓的H函数，它随着时间而减小。这个函数为熵的箭头提供了一个互补箭头，熵在系统向平衡态演化过程中是随时间增加的（H函数在数值上与熵相等，但符号相反）。因此玻耳兹曼声称，他在分子层次上解决了不可逆性佯谬。

这对于玻耳兹曼来说是一个极其重要的成就，因为他一直持有这样的观念：“力学是整个理论物理学大厦的基础，是其他一切科学分支的根基。”但是，为了得出他的不可逆方程，玻耳兹曼做了一个关键近似——即“分子混沌”假设。换句话说，他认为分子在快要碰撞之前是彼此不相关的，但是在碰撞之后它们就变得彼此相关了（因为它们的轨道由于碰撞而发生了改变）。因为这个分子混沌假设是时间不对称的，这就解释了为什么玻耳兹曼方程描述了不可逆的时间演化。

因而，非常自然地，分子混沌破坏了牛顿的时间对称的定律。正因为如此以及其他一些原因，玻耳兹曼受到普朗克年轻的助手泽梅洛（Ernst Zermelo）以及他自己的朋友罗史密茨的尖锐批评。用更直截了当的话来说，他们认为，不可能希望在无视任何时间方向的方程中，建立起一个唯一的时间箭头。罗史密茨1876年的批评，所根据的是力学的时间对称本性。20年后，泽梅洛对玻耳兹曼的批评，则是根据庞加莱的回归论，我们可以回忆一下，它讲的是每一个孤立系统迟早会回到它的初始状态。

这些非难使玻耳兹曼如此震惊，他想从通过力学找到时间箭头的最初打算后退。以致他最后认定，第二定律根据的是几率理论而不仅仅是力学，因而必须摈弃力学归并论，而采用原子论的形式，并且原子也必须用统计方法处理。他后来甚至于赞同这样的观点，即宇宙早就已经处于热寂的状态，或者说是总体热力学平衡的状态，但是我们恰巧位于一个有涨落的区域，涨落使我们离开平衡态，现在它又使我们朝着平衡态回返，同时伴随着熵的增加。这样，玻耳兹曼的主张就是，在宇宙尺度上是没有时间箭头的，有许许多多和我们时间箭头相同的区域，也有同样多的区域，在那里时间箭头和我们的相反。

这是一个有独创性的想法。不幸的是，它是站不住脚的，因为我们从来没有观测到过其他的宇宙部分，在那里时间是颠倒过来的，例如一头公牛越长越小，并且可以把它糟蹋了的瓷器店还原一新。确实，正如我们在第三章中提到过的，现代天文学和宇宙学表明宇宙是

在膨胀着的，所以它不可能处在热力学平衡状态。这里必须有一个宇宙学的时间箭头，因为在较早时期，宇宙中的星系是聚在一起的。所有已知的现象都与一个单一方向的时间一致，此时热力学平衡态是在将来而不是在过去。玻耳兹曼最初想使热力学的不可逆性与量子力学和经典力学的可逆性相一致，他的目的是对的。他走在了他的时代的前面，但是缺乏能够战胜他论敌的知识。我们将在第八章再回到这个问题。

在生命的最后几年里，玻耳兹曼仍然热中于论证热力学行为是原子和分子现象的表现形式。如我们以前提到过的（第一章），这引起了他和他的对手马赫以及奥斯特瓦尔德之间的激烈论战，并且无疑地，由于玻耳兹曼说话啰唆，加之偶尔不加说明地改变自己的观点，这使得论战常常火上浇油。玻耳兹曼的对手是有相当权威的人，又加上玻耳兹曼在说话时声音不够洪亮，这使得他在论战中常处于下风，使他感到沮丧，所以在来比锡时他就有了自杀的念头。当然也不是说他的对手总是占尽上风。德国数学物理学家索末菲，这样描写过1885年在卢比克科学会议上的一次争论：“玻耳兹曼和奥斯特瓦尔德之间的论战，像是一头公牛和一位灵巧敏捷的击剑手之间的争斗。但是这一次不论这位剑手的技艺多么高超，公牛还是赢了。我们年轻的数学家们都是站在玻耳兹曼这一边的……”然而大多数情况下，玻耳兹曼进行的是孤独而艰难的战斗。在新的一轮反对浪潮的冲击下，玻耳兹曼在他的经典著作《气体理论讲义》的前言中写道：“我意识到自己是在孤零零地奋斗，对时代潮流做软弱无力的反抗。”

玻耳兹曼在1902年从莱比锡返回到他在维也纳的职位。此时他是创立动力学理论的三巨头中唯一的幸存者——克劳修斯和麦克斯韦都已然去世。玻耳兹曼一方面在学术上孤立，另一方面健康日益恶化——他的视力越来越差，常常哮喘、心绞痛和头疼。1903年他的妻子汉丽蒂（她把他叫做“可爱的胖大令”）写给他们的女儿埃达说：“你父亲的身体每况愈下，我已经对他的将来失去信心。我曾幻想过在维也

纳的生活会更好。”

在1904年庆祝他的六十大寿的生日宴会之后，玻耳兹曼变得更加悒悒不乐。在1905年到1906年之间的冬季，他讲完最后一节理论物理课后，一位学生这样描述过玻耳兹曼的痛苦：“一种神经质的病（头疼）使他不能继续教学。我和另外一位学生一起，在他的别墅里通过了口试。考试结束后离开时，我们在前厅听到了他令人心碎的呻吟。”一种害怕失去创造性的先天性恐惧以及心绞痛的恶化，加上攻击性的论战，终于使玻耳兹曼在1906年9月5日自杀了。仅仅几年之后，他的原子论学说开始流行起来，这在相当大的程度上是由于爱因斯坦关于布朗运动和分子计数的研究以及佩林（Jean Perrin）很快在苏黎世对此所做的实验证实。在约翰斯顿（William Johnston）名为《奥地利精神》的一本内容丰富的书里，他谈到了玻耳兹曼所处时代的奥地利知识界，自杀几乎成了一种时髦的风尚，以此来逃避动荡年代的坎坷人生。玻耳兹曼确实有一次说过，“支配我全部心思和行动的，是发展理论”，“为此我可以牺牲一切：因为理论是我整个生命的内容”。

## 一切都在脑际：“粗粒化”

玻耳兹曼方程现在仍然生机勃勃、充满活力。它被广泛用于描述不可逆的“输运过程”，例如像稀薄流体混合物中的扩散和黏滞性，此时玻耳兹曼的假设实际上是一个很好的近似。也有一些人尝试过改进H定理。当初最有希望的是所谓粗粒化方法。不幸的是，它的结果却导致与玻耳兹曼的看法完全不相容的结论。这结论认为时间箭头是纯主观的，只是通过我们所用的近似才存在于微观世界之外。

粗粒化是一种技巧，用于描述我们不能直接观察到的较小尺度上的事件。它可以用来计算一个系统（例如某种气体样品）任意限定的空间亚单元中，分子的平均运动。运用这种处理方法，我们可以有效

地挽救玻耳兹曼的H定理（它等价于熵的增加）以及与此有关的系统演化的“不可逆性”。这样，在微观的可逆方程与时间箭头之间就建立起联系。

但是，没有什么东西可以告诉我们，粗粒化的程度到底应该如何。用这种方法计算出来的熵，决定于所选取的粗粒大小，这直接与热力学发生矛盾，因为热力学中。熵的改变完全是客观的。除此之外，如果用事后决定的方式进行粗粒化，则不能保证熵会随时间增加——也许它会随时间减少。即使粗粒化的方法可行，也不过是因为我们在某一尺度上引入了近似，而忽略了在更小的尺度之下发生些什么。这使得我们处于一种奇特的情况，即可以任意找一个截止点，用来结束可逆的微观世界，开始不可逆的宏观世界。换句话说，只有在这一个主观决定的界线之上，时间才具有方向的意义：粗粒化方法把整个不可逆性和时间箭头的问题，降格为一种幻术。

如此一来，时间箭头再次变成一种主观的概念。我们为了追循一个热力学系统的行为而做出的近似，被说成是因为我们不可能追循构成系统的亿万个分子的运动。如物理学家杰恩斯（Ed Jaynes）所说：“不是因为物理过程本身不可逆，而是因为我们追循物理过程的能力有限。”这就是说，如果我们的感觉足够灵敏，我们就可以看到分子的单独运动（忽略这在量子力学中引起的问题），从而可以去证实，所有的过程在这种微观层次上真正是可逆的。这马上会使熵的概念，变成我们对一个过程精确细节忽略程度的量度。这个想法在一个叫做信息论的学科里得到了成功的发展。

信息论处理的是译码和发送信息的问题。任何一个信息交换系统——从高保真度收录机到电脑或电话——都无法摆脱随机性的干扰即噪音的影响。信息论的目的就是：从伴随着噪音而接收下来的一切信号中，提取真正有用的信息。这一理论的基础是山侬（Claude Shannon）和魏沃（Warren Weaver）在1949年奠定的。就他们所考虑的来说，信息可以由一串毫无意义、杂乱无章的信号组成。信息在

技术上唯一的重要性，就是它可以被编码、传送、选择和解读。山侬提出了一个信息的纯数学定义，它可以用于一个系统中任何几率分布的情况。利用这个定义，可以计算出在被干扰得一塌糊涂的情况下，发现信息的几率。

山侬的数学公式看上去很像统计力学中熵的公式。许多人认为，这表明这两个概念之间有一种直接的联系，于是这就产生了一个结果，即所谓的“最大熵”技术。这一技术极其具有威力，它可以使我们从一大堆乱七八糟的干扰噪音中，把一丁点有用的信息找出来。噪音的本质是随机、无序的，这与信息的有序性形成对照。可以说，信号中的信息量越大，它的熵就越小。按照杰恩斯和他的助手的主观性解释，信息“熵”是一种量度，它表示忽视观测尺度之下过程细节的程度——这相当于可逆过程与不可逆过程之间的截止点。信息本身定义为山侬的熵的负值，所以有时也称为“负熵”。

信息论中包含了丰富多彩的内容，它在计算机理论和通讯系统工程问题的分析中，有着极其广泛的应用。然而，熵和信息之间的极为相似，并不意味着它们必须都是主观性的概念。事实上，作为信息论的开山鼻祖，山侬是在数学家纽曼的劝说之下才把“熵”这个字眼引进他的讨论中的，据说纽曼曾对山侬讲：“这样做会使你在辩论中大占优势，因为没有人真正了解熵到底是什么东西！”

不应当认为信息论是为粗粒化打抱不平，而且是向主观主义倒退。事实上，在最大熵技术中并没有用到粗粒化——它并不是做某种任意性的分割，而是把整个系统作为一个“黑盒子”，其中细节我们是不知道的。不论我们从其他什么地方发现一星半点对它的支持，粗粒化已是无可补救。如普里高津和斯坦格斯所写道的：“不可逆性或者对所有的层次都对，或者对所有的层次都不对：它不可能在从一个层次过渡到另外一个层次中间，无缘无故地突然冒出来。”我们在这一章中已经谈到远离平衡态时发生的不可逆过程，在下一章中我们将更详细地讨论这个问题。特别是，在生命本身的存在和维持所必需的、

许多关键性的生物过程中，不可逆过程起着基本的作用。如果粗粒化或者主观信息论的说法是正确的，我们则不得不接受这样的观念，即所有这一切皆为虚幻。确实，它会给出一个佯谬结果，也就是说，像我们大脑的功能一样明显是不可逆的宏观过程，都仅仅是由于我们所用的近似所致。

## 宇宙学的时间箭头

我们已经看到，第二定律如何意味着宇宙的热寂，也就是宇宙最后演化到一种彻底无序的状态。这种看法和宇宙学家们的看法是否一致呢？

在前面几章里，我们已经简略地谈到过时间和宇宙的起源。我们知道宇宙是在膨胀的，而且我们可以预言它的两种可能的极端命运：继续膨胀直到热寂（虽然也有人推测，如我们将在第八章看到的，即使在这样的条件下，也会出现一个恢复了活力的宇宙）；或者是大坍缩，此时无处不在的引力最终使膨胀停止，并且使所有的物质不可抗拒地回聚到一起，从而形成一个最终的奇点。即使两者之一中有一个是对的，我们现在也无从得知究竟是哪一个对，因为这实际上决定于宇宙中现有物质的数量，这一点我们在第三章中已经提及。

假设宇宙是闭合的而且在坍缩。给定熵的增长和时间箭头之间的关系，是否大坍缩就意味着，一旦坍缩开始，时间就会逆转？有些人认为是这样的。河水将会倒流。布里斯托尔大学的贝里（Mike Berry）对此讥讽到："光线会从眼睛里发射出去然后被星辰所吸收。"在这些离奇的想象背后，是这样一个观点：在膨胀过程中，时间箭头是从高度有序的大爆炸奇点，指向某种无序性最大的中间态；然后当宇宙开始向看上去和大爆炸同样高度有序的大坍缩收缩时，时间箭头便反转过来。彭罗斯据理驳斥了这种观点，他认为，即使在大坍缩的过程中，熵也还是增加的，第二定律仍然有效，时间箭头也保

持不变（虽然他仍然把第二定律看做是自然界的“二级”定律而不是“一级”定律）。这是由于，大爆炸和大坍缩这两个奇点的结构是不等价的。许多宇宙学家赞同一种对称模型，即大爆炸和大坍缩是不可区分的，因为两者都是物质无限压缩的火球。然而，彭罗斯认为，所有的原初时空奇点都具有一个限制条件，它并不适用于黑洞或者最终的奇点：大爆炸奇点相对于大坍缩来说，有序程度要高得多，熵也要低得多。这个令人惊奇的结果是由于时空在奇点附近的几何结构，它对于大爆炸和大坍缩是不同的（图17）。观测证据表明大爆炸奇点是各向同性的——像一块牛奶冻，不管从什么地方把它切去一半，它显不出有任何结构——并具有高度的有序性和低的熵。但是在走向大坍缩的过程中，会产生像黑洞那样的时空缺陷。它们在大坍缩中凝聚成质量巨大的乱糟糟的一团，具有像果仁蛋糕那样的无序结构和相应的高熵。如果原初奇点没有这样的限制，就不会有第二定律，而且我们也就会期待像发现黑洞那样发现白洞。

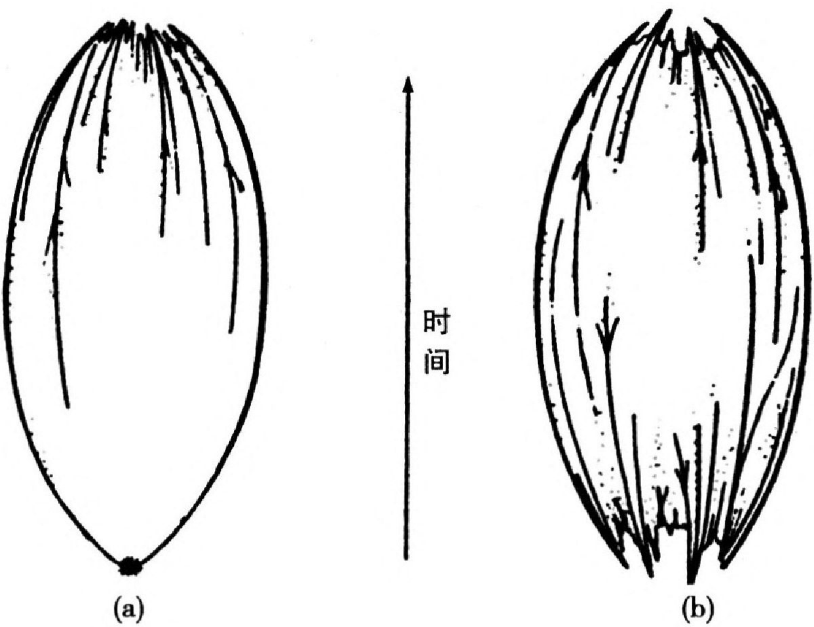



图17　（a）一个闭合宇宙的历史，它开始于受到高度约束的低熵大爆炸，而终止于混乱的高熵大坍缩。（b）如果没有特殊的初始限制，大爆炸就同样是高熵的。在彭罗斯的图像下，只

有（a）才能在宇宙学尺度上满足热力学第二定律。[录自彭罗斯所著《皇帝新脑》第339、第341页。]

如果彭罗斯的猜想是正确的，则我们需要知道，为什么在这些奇点中有这样的时间不对称性，使得产生低熵的大爆炸和高熵的大坍缩。许多物理学家也许简单地满足于这样的看法，即特殊的低熵大爆炸状态仅仅是一个“初始条件”（出自上帝之手？）的结果，如此而已。然而彭罗斯却认为，在时间“开始”和“结束”时奇点的独特性质，显然表明量子引力的整个理论必须是时间不对称的。按照他的看法，一个完全令人满意的理论，应当同时对时间演化和初始条件做出解释。到目前为止，看来我们离这个目标还很遥远。

尽管如此，彭罗斯的想法看来已经给出了时间箭头的正确条件。还有一个问题没有解决：如何安排时间的流逝，使其由低熵态——大爆炸，流向高熵态——大坍缩。用现有的与时间方向无关的物理规律，在这两个极端状态之间画上一个箭头，殊非易事：为什么不从大坍缩开始向大爆炸走呢？彭罗斯的说法根据的是用粗粒化来计算熵，这里面就包含了主观主义的各种问题。他承认不同的粗粒化会给出不同的结果，但他认为这在实际上不会造成很大差别，因为在开始和结束时刻所涉及的这两个熵值是有“天壤之别”的。

彭罗斯的推测还有另外一个有趣的推论。他认为一个完备的、具有时间箭头的量子引力理论，可能会解决第四章中讨论过的现代量子论中的一个中心问题——即如何理解测量过程。一个包括时间箭头的量子引力理论，也许能够描述不可逆的波函数坍缩，只要存在足够大的时空曲率。事实上，引力相互作用会使波函数砰然爆裂，这样就提供了一种解释，为什么坍缩只能在宏观尺度上发生（由于在这样的条件下有大量的粒子存在，因而就会有可观的引力作用）。彭罗斯承认说：“到目前为止，对于我认为大有需要的新理论来讲，这只能说是刚刚有了一个萌芽。我相信，任何完全令人满意的新理论，必须含有关于时空几何本质的某种根本性的新思想。”

最后，我们注意到，高度有序的大爆炸这个观点使得人择原理（见第三章）看来有些靠不住了。人类在宇宙中出现的机会可能是非常小，但是人择原理认为，我们能够在此提出这个问题，本身就表明命运对人类的创生十分垂青。然而彭罗斯低熵大爆炸的初始条件，比起人类的创生来，其实现的机会真正可以说是小到分子是一，而分母是个天文数字。无疑地，极端形式的人择原理——即宇宙是为了有利于人类而创生的——已经站不住脚了，因为人类存在几率之小远不及宇宙存在几率之小。

虽然我们现在对于熵的意义有了一个相当深刻的概念，但是我们仍然没有解决微观世界和宏观世界之间的矛盾。热力学和力学之间的冲突要到第八章才可以解决。在此之前我们想表明，为什么热力学的内涵是如此重要和广泛，不能草率地把它当做附加在力学之上的一种主观主义的旁门左道。为此，我们将在物理学和整个生物学中，浏览一下不可逆性问题。

# 第六章　创造性演化

化学反应是含有时间之箭的一大类过程：这些过程都是不可逆的。有些化学反应显示着非常明确、非常有规则的变化——它们简直就是一种不折不扣的化学“钟”。对这些过程进一步研究，会让我们了解到，它们里面的时间是如何以及是为什么会滴答消失的。进一步说，既然有机细胞的化学性质就是生命的精髓所在，这些钟也就是转动我们自己身体中诸齿轮的微观“轮齿”。时间之箭将以“生命之箭”的方式出现。在这出现的过程中，它产生的花样是如此细致，如此丰富，使我们实在不能相信它是“简并派”人们所说的幻觉。

内禀于第二定律之中的时间之箭，并不等于一直走向无序的盲目破坏；相反的，从第五章开始讲的，支配远离平衡的不可逆过程，对其原理的研究，能帮助我们了解，缓慢而无情的衰退、错综复杂的生命图样、湍流的泛滥，这三者之间的界线如何划分。

一个系统只有被驱赶到远离平衡的状态之后，才能开始产生我们感到兴趣的节奏。试以一瓶啤酒做个有趣的比方。瓶子直放在桌上时，啤酒处于力学平衡。斯文的饮酒者将酒瓶稍加倾斜，使酒平稳地流入杯中，这样他把啤酒移到一个“近平衡”的稳态。可是一个急于解渴的人，抓起瓶子就喝，很可能就会把酒推到远离平衡的状态。他会发现，如果瓶子斜过一定程度，啤酒便会很规则地汩汩流出来。对酒瓶来说，存在一个“临界角度”，酒瓶一达到那个角度，酒就开始来回摆动。那个角度是一个转折点，是一个从混乱到组织的跳板。

这当然就是上章检查自组织的热力学基本配方时，开始考虑的课题。将第二定律应用于任何一个开放系统，即一个可以输入也可以输

出物质和能量的系统。将此系统从平衡状态远推到一个转折点，组织便可能出现。我们已经遇到过这样的例子：一个化学反应达到它的转折点以后，出现有规则的颜色摆动，形成为一台十足的化学“钟”。我们目前的任务是，当像这样的钟远远离开它的第一个转折点以后，我们应该用什么方式来描述它。

光靠热力学是不够的。热力学通过熵增的倾向描述时间之箭，它只在去往平衡的路上，放置了指路牌，只告诉我们什么地方变化会发生。但至于会是何种变化，它却给不出任何线索。热力学中没有一个万能法则，告诉我们一个系统如何在时间上演化。我们不得不向热力学告别，而开始跟一些崭新的技巧打交道。

## 秩序和混沌

有人或许会想试用量子物理学或经典物理学来描述化学钟。这个办法将是极端复杂的，但即使不管这一点，我们也得弃之不用，因为这两套理论都不区别时间的两个可能方向。我们必须另想办法。想知道像火车的往来那种日常的事情，我们查一下时刻表就行了，无须了解火车运行的方式，不必管火车是用蒸汽的，还是用电的，还是用柴油机的。类似地，要描述一台化学钟，我们搞一套纯经验性的报道就行了，这项报道不必等分子层次完全了解以后就可以写。

这就是所谓现象派的办法；当不可逆过程推到极点时，这种办法特别有用。再以我们那瓶啤酒为例。要是我们把酒瓶整个倒过来，那就会出现比斜着的时候的咕噜咕噜还更妙的行为：啤酒的流出将变为湍流式，出现涡卷，不计其数四面乱跑的分子组织成漩涡。同样地，我们也将看到像化学钟反应那种不可逆过程，如果超过某个极限，混沌便会发生。这就是上章提到的动力学混沌：这里，严密规律所产生的行为，看上去是随机性的，其实是具有很细微的组织。有些科学家相信混沌支配着各种复杂的现象，例如病人心脏不规则的跳动，野生

动物总数乍看上去毫无规律的涨落，气候在时间上的变化，等等。

初看上去，混沌似乎跟化学钟的有组织行为迥然不同，其实两者在物理上（数学上）系出同门。这一点很重要，其基础就是时间之箭。假设时间是连续的，而不是一系列分离的时刻，所有的耗散系统便可以用“微分方程”来模拟。“微分方程”是瞬时变化的数学描述。北美洲吉卜赛蛾总数的变化率也好，化学钟各种含量的变化也好，微分方程都可以同样应付。这些方程和牛顿方程、爱因斯坦方程、薛定谔方程不同，这些方程本身具有时间之箭。它们允许各式各样众多的解——从自组织到混沌——它们从多方面说明，为什么我们这世界是如此丰富多彩。对于它们在生物学中的含义，我们将在本章和下章中加以探讨。

这些时间不对称的微分方程里面，是什么因素使它们既可以产生秩序，又可以产生混沌呢？是“非线性”。前面讲热力学时我们已经注意到，“非线性”的意思就是：“所得非所望。”一个线性关系中的量是成比例的：10个橘子的价钱是1个的10倍。非线性意味着批发价格是不成比例的：一大箱橘子的价钱比1个的价钱乘橘子的个数要少。这里重要的观念是“反馈”——折扣的大小倒过来又影响顾客购买的数量。

互为因果，听上去很简单，但常会引起出人意料的现象。它能使系统变为不稳定，使它达到一个临界点，就像麦克风和喇叭之间的正反馈，把悄悄耳语通过一个放大回路弄成震耳欲聋的狂吼一样。在化学钟里，当一种化学剂的产生影响它下次产生率时，反馈就出现了。是反馈把化学混合体中的起伏放大成化学钟里的前后呼应的颜色变化。也就是反馈使旅鼠群落过早死亡：旅鼠可能繁殖太快，把能吃的食物一下子都吃光了。生物化学界也有类似的情况，例如，一个反应产生某种酵母，该酵母的出现又鼓励自身的生产，终于把所有的反应剂都耗尽。在生物界中，各式各样的正反馈和负反馈凑合在一起，把细胞核酸DNA中的遗传蓝本发展成复杂的有机体。

只是最近20年来，我们才开始开发耗散式非线性方程中的潜力。线性方程可以用已知的数学解析手段来研究、来解，而非线性方程，除掉少许特殊情况之外，很难用这种方法来解。这里只好用最老实的办法——把数字输进方程，一步一步用数字计算求解。这就是多年来潜能如此丰富的这项领域很少人问津的主要原因：计算机问世以前，这种工作无法进行。但现在我们可以详细地探讨这座非线性原始丛林了。

非线性数学看上去很“邪门”。在计算机还未成为日常工具以前，人们采用近似方法粗略描述非线性系统——把系统线性化。试估计一下加薪以后要付的税。虽然一般说来税收规则是非线性的，和个人收入的关系很复杂，可是为了得一个粗略数目，我们不妨假设税只照“起码率”付。但这种线性手段，用途迟早是有限的，不仅是处理一年度的税收申报，处理非线性动力学也是一样。贫血的线性近似得不出来新花样，只有新的，非线性的变化才可以产生我们想解释的各式各样的组织和混沌。下面我们将看到：运用非线性动力学，加上一种动力学行为的生动描述——所谓“分叉分析”（见第162页），我们可以对远离平衡的时间演化，进行比只用热动力学时详细得多的调查。

## 培养自组织

最简单的过程中可以出现意料不到的非线性效应。上章我们见到热扩散的例子（图15）：当两种气体的混合体由于加热而离开平衡态后，组织便会以一种简单浓度梯度的形式出现。那里，离平衡只少许的偏离便可导致宏观的秩序。然而此种现象，如果跟当系统离开它第一临界点以后，自发涌出的壮观的组织比起来，仍是小巫见大巫。夹在两片玻璃之间的一薄层液体，对它加热就可以使组织出现，形式是六角形对流单元组成的蜂巢结构。这对任何一个因循传统的、基于平衡的世界观的人来说，是一件很惊奇的现象。我们总以为加热越多，液体里的分子便会跑得越快、越乱，怎么反而出现结构了呢？

蜂巢状的自组织是法国科研者贝纳（Henri Benard）于1900年首次发现的。1916年瑞利男爵（Lord Rayleigh）曾尝试对其进行解释。现在我们知道它是来于所谓“瑞利-贝纳流体力学不稳定性”（参见彩色插图）。

实验是把液体盛在一个透明碟子里，把碟子放在像烧菜用的电炉一类的热源上。热以传导或对流或两者兼有的方式从碟子的底部升到顶部。在加热以前，液体看上去是平静的，尽管微观层次上是分子在作或多或少的随机运动。一旦加热，液体在垂直方向便产生一个温度差。然而液体在宏观层次上依然是静止的，直到顶部和底部之间的温度差达到某个阈值之后，情况才会改变。阈值未达到以前，热的输送只是以传导的方式进行。超过阈值以后，对流开始，底下较热的液体流入上面较冷的部分。在此同时，蜂巢图案由于浮力、热扩散、黏滞力三者之间的耦合而产生。

如果按照基于平衡的想法，我们便会以为加热越多，不计其数的分子便越是在碟子里到处乱窜。可是，看一下彩色插图就知道，蜂巢结构是远比加热以前的情况更有组织。蜂巢结构的尺度是个别分子之间距离的1亿倍。为了形成这个蜂巢状的对流单元，无数分子必须在如此巨大的尺度上“齐步运行”。只要把温度差保持着，这个蜂巢结构肉眼就看得见。热的耗散把熵从系统中输出，而使蜂巢结构维持下去。

如用热动力学的描述，六角单元出现的温度就是上章讲到的临界点或分叉点。在该点，系统有两条路可走。例如在瑞利-贝纳不稳定性的情况下，邻近单元具有相反方向的对流运动。温度一旦超过临界值，这些单元便肯定出现，这是毫无疑问的。可是单元旋转的方向是不能预测的，它是每次实验微观层次里，许多控制不了的涨落升级到宏观层次的结果。普里高津的同事尼古力斯（Gregoire Nicolis）说，该现象是出于“偶然性和决定性之间一件出色的合作”。

要培养出自组织，涨落是必需的种子。在离平衡态不太远的地方，液体中的对流很小，很守规矩，作用不大。这时涨落像临死的人说的耳语，很快地就消失了。但是如果存在有反馈，这句耳语就会变成狂吼。在远离平衡的状态之下，系统的各种非线性性质将把微观对流放大成覆盖整个碟子的组织，形成一个液体的蜂巢。有些人想用热动平衡的语言来解释这个现象。但这种语言对比方说冰晶那种单调的规则重复还可以应付，要它来描述像瑞利-贝纳细胞组织那种“非静止的”、耗散式的结构，是完全做不到的；这个组织只是液体在加温状态之下才能存在。

大多数的化学家和分子生物学家对世界的看法，是着重个别分子的活动。这种办法对许多处于平衡状态的系统来说是很有效的。可是它表达不出一个具有自组织的介质里的分子和分子之间的“信息交流”。处于平衡状态的冰，水分子间彼此影响的作用范围不超过一亿分之一米，而出现在耗散式系统的结构相对庞大得多，大到厘米的量级——这种动态结构类似成千上万的冰箱以同样的速度制造冰块，类似全纽约的居民同时在做同一个体操动作。

然而自组织并没有什么玄奥；下面将会说明，它是在远离平衡态的情况下，含有时间之箭的物理定律的必然后果，尽管这是个出人意料的后果。我们的兴趣并不是在“时间的尽头”平衡热力学统治一切时，某个个别化学反应的去向；我们感到兴趣的是，在到达平衡态的途中，亿万个分子居然会如此步伐协调，在空间形成宏观的图案，在时间上出现大规模的振荡。

要知道这是何等出人意料之外的事，设想一辆载满网球的卡车，一半网球是白的，一半是黑的，均匀地混在一起。对液体加热就相当于把卡车开在高低不平的道路上，使网球剧烈地相互碰撞。设想网球在这沸腾状的场合中，居然排出了一个规则的图案，好比说，所有黑网球都跑到车子的一端，所有白网球都跑到另一端。瑞利-贝纳细胞组织所显示的大规模秩序是同样地令人注目：它意味着巨大数目的个别

分子在时间和空间上的同步行动。真正比起来，后者是更令人惊讶的，这些分子形成的结构，相对来说，远超过网球的图案。看上去好像是，在离平衡态足够远的场合，每个分子都有同一个时间感，都按照那个时间齐步动作。整个系统变活了——它不能再被看为一群四面乱跑的分子。液体中的这些分子自发地自我组织起来了。

## 化学中的自组织

要把能起化学反应的混合物保持远离平衡，是很简单的。只要把它放入一个不断搅动的流水式反应器就行了，而这种反应器是每个化学工厂的典型设备。化学物品进入反应器上方，由于搅动而起化学反应，成品再从低处拿走。如果某个成品（叫它X）催化自身的生产——一种所谓“自催化”的反馈，各种非线性的种子便种下了。X在某个时刻形成的数量要看X当时有多少，这种非线性的特征类似麦克风和扩音机之间的正反馈。这样，并不太麻烦，我们就具有了自组织和像化学钟那类现象所需要的各种因素。

第一个考虑到这种可能性的是数学家杜灵（Alan　Tur-ing），他是20世纪英国科学界最伟大的人物之一。他这一套想法写在1952年《皇家学会哲学丛刊B部》发表的一篇出色论文里面，当时他40岁，在曼彻斯特大学工作。智力成就点缀着杜灵的一生：他本人就像座人体化学钟，每五年鸣响一次，这是最后一次。1935年，他创造了“通用机器”的概念，把一个描述精神的简单机械图像和纯数学结合起来，说明机器可以模拟思想；1940年，他在白金汉郡布莱其列镇当密码专家，他的计算才能被用于破译德国海军情报，情报先用一台“厄尼格摩（谜）机”编码；1945年，他从事制造“自动计算引擎（Automatic　Computing　Engine）”，这是他的通用机器也就是电脑的实际体现；最后，在1950年，一生许多时间花在破译敌人密码的杜灵，将他的注意力转向大自然用以产生各种图案的密码。

杜灵当时的兴趣是想为形状、结构、功能在生物体中的出现，即生物学中所谓的“形态来源论”，找出一个化学基础。杜灵问他自己一个简单的问题：一个有机体是如何把一个化学浑汤整理成为一个生物结构，如何使一团一模一样的细胞变成一个有机体？这是生命最大的难题之一。然而他在论文的提要中写道：“本理论并不提出任何新的假设；它只是说某些熟知的物理定律就足够解释许多事实。要全部了解本论文需要相当程度的数学，需要少许生物学和初等化学。”

让我们来考虑一下胚囊的发展过程：一个哺乳动物的胚胎，本来是一个许多细胞组成的球体，这球体逐渐失去它的对称性，有些细胞发展成头，有些发展成尾巴。从一个完美的球体开始，我们也许会以为，支配它发展的生物化学反应的均匀、不可逆的扩散，会保持这球对称性，那么我们每个人都应该是一团一团的球体了。然而在这篇论文里，杜灵证明，受精卵变成生物复杂形态所必需的这种对称性的破坏，的确可以出现。他这里的思想定性地说，是和我们上面已经讲过的一样：在平衡态附近，最对称的均匀状态是稳定的；远离平衡态，均匀状态就会因为到处存在的涨落而变成不稳定了。杜灵打了个机械式比喻：“一根棍子如果从它引力中心稍上的一点吊着，棍子将是处于稳定平衡。但如果一个老鼠沿着棍子向上跑，平衡便迟早会变为不稳定，棍子便会开始摆动。”

卵很少是球对称的，而且诸如引力的因素会破坏这对称性。尽管如此，从杜灵的想法可以很生动地描述出自然界各种图案的产生，蜗牛壳也好，蛇皮也好，这些都将在下章详述。这里的过程当然都是不可逆的，都含有时间之箭。可悲的是，杜灵没有能更多地发展这套思想，他讨论形态形成的论文发表两年后，就自杀了。

英国社会50年代的道德风气使杜灵活不下去。1952年，杜灵的末日开始来到，他以“违反1885年刑法修正法第二节，犯粗鄙行动罪”的罪名，被提入法庭受审。先是由于警察调查他家发生的盗窃案，使他自招是个同性恋者。他被处缓刑，送入医院接受医药治疗，

注射了降低性欲的荷尔蒙。可是在1954年圣灵降临节的星期一，50年来最冷最湿的一天，杜灵吞食了泡过氰化物的苹果。郝基斯（Andrew Hodges）在他写的《杜灵传》里讲道，1939年杜灵在剑桥看了《白雪公主》的电影："他非常欣赏恶巫用绳子吊着苹果，晃来晃去放进沸腾的毒药锅里，一面口中念念有词：'苹果泡呀，泡呀，睡觉一般的死亡，泡进去吧。'这几句话，他本人就喜欢念来念去，日后居然应验了。"

杜灵的死一如玻尔兹曼的自杀，是科学界的一大创伤。幸好他已经取得一项重要的发现，他发现如果多种颜色的物体具有不同的扩散率，在液体里相互起反应，它们便会变化其浓度而形成空间的图案。这个现象是违背直觉的，因为我们总觉得任何不可逆的混合过程，结果总是把原有的图案、结构洗刷一尽，就像咖啡加牛奶而产生的花样迟早总要消失一样。杜灵远远超过他的时代，他写出了数学配方，既可以制造不随时间变化的稳态图案，也可以取得像化学钟里的彩色波浪的振荡式图案。现在我们把离开平衡态足够远的、图案首次出现的那一点，叫做"杜灵不稳定性"点，这是我们上面提过的"临界点"的一个例子。可是，在该点认为应该出现的图案，虽然理论上可能，当时还没有在任何一个实际化学系统里切实地看到，并且他的模型里的有些细节由于其他原因，也受到过批评。尽管如此，生物界里自组织是很普遍的，其中有些例子是可以用"反应-扩散"理论来解释。但目前我们先把注意力集中在一些比较简单的化学现象上。

此后15~20年，杜灵的工作可说没有受到化学家和生物学家的注意。其原因很多。为了处理所涉及的非线性方程，杜灵采取了线性化的办法，即假设在有一定限制的情况之下，数学行为是线性的，是可预知的。这样一来，他的分析就难免太"近视"了，离开平衡态以后，不能超越过第一个临界点。换言之，杜灵可以说什么时候将有图案出现，但他不知道，当系统继续远离平衡态时，那幅图案将会起什么变化。杜灵意识到要继续发展这项工作必须要用高速计算机，而这样的

计算机当时是没有的。再说，当时还没有任何人知道有什么化学反应，是杜灵的理论可以应用的。

## 布鲁塞尔振子的诞生

目前许多实验室在从事自组织的研究，用的方法和杜灵原来用的大致相同。近20年来，特别是两项关键性的发展，大大提高了人们对这方面的兴趣。一项是1968年在布拉格举行的讨论会上，西方的科学家首次听到魔术似的“贝鲁索夫-扎孛廷斯基化学反应”（下面很快就要详述），并且把它和生物界发生的一些振荡加以比较，这些振荡帮助生物利用能量，例如酵解和光合作用。另一项发展是普里高津和勒菲弗（Rene Le-fever）的工作，也是在1968年发表的。他们引用了杜灵的启发性论文，构造分析了一个具有空间自组织必需条件的、起化学反应的模型系统。1973年，弗吉尼亚工艺大学的泰森（John Tyson）给这个模型命名为“布鲁塞尔振子”，因为它诞生在比利时首都。在这篇论文里，普里高津和勒菲弗证明了布鲁塞尔振子出现的方式符合本书第五章提及的“格兰斯多夫-普里高津热力学演化准则”。（这里重提一下：该准则基于热力学第二定律在远离平衡的场合的运用。）这是他们能找到的模型之中，既满足可能出现热力学不稳定性的演化标准的条件，又容易处理的最简单的模型。因此这模型具有坚固的热力学基础。此后，勒菲弗和尼古力斯证明了布鲁塞尔振子可以在某些化学物质的浓度上，显示出持续不断的、规则性的振荡。

在诸如布鲁塞尔振子的化学钟的情况下，不难看出，反馈和非线性对自组织是少不了的；这里，巨大数目的分子在产生图案的过程中，似乎在互通信息。布鲁塞尔振子是一个理想化的模型，它牵涉到两种化学物质A和B，它们转化为另两种物质C和D。为了产生有趣的非线性现象，转化不是在单独一个化学反应里完成，而是由四个基本步骤组成，其中牵涉到两个中介体X和Y。详细情况并不复杂：一个A分子先转化成一个X分子，这个X分子在第二步和一个B分子起作用，

产生Y和C。第三步是两个X分子和一个Y结合产生三个X分子。最后的反应是X直接转化成D。到达自组织的“跳板”——非线性反馈，出现在第三步，那里，从两个X分子，经过和中介体Y的反应，得出三个X分子。整个过程于是存在有反馈，因为其中一个分子牵涉到自身的生产，它“自催化”了。于是非线性出现了，因为每两个起反应的X分子，都要产生另一个，一共变成三个了。

如果化学原料未用完之前再加满，布鲁塞尔振子就可以保持在远离平衡态的状态。为此，只要把反应放在一个不断搅动的、开放式的反应器中进行。我们控制着A和B流入的速率，使它们保持适当的浓度，同样地，我们也控制着产品C和D的浓度，唯一随时间变的就是X和Y的浓度。如要知道它们是怎么变，就得写出布鲁塞尔振子的数学描述，这是一系列有关X和Y的耦合微分方程，并且对这些方程求解。

隐藏在耦合微分方程里面的错综行为，可以从布鲁塞尔振子的行为中窥见一斑。这里的数学分析相当复杂，但如果设想X是红的，Y是蓝的，结果就可以叙述如下：

让我们先考虑化学反应完结时的情况。各种成分混在一起，互起作用，反应达到平衡，化学变化停止。在平衡态下，是一种普普通通的紫色的液体，一种红分子和蓝分子的混合体。如果A和B的浓度保持在平衡值附近，保持在普里高津最小产熵定理有效的定态区以内，就不会发生什么大变化。只是当A和B的流人率超过平衡浓度以外某个阈值以后，有趣的现象才会发生。（在啤酒从酒瓶流出的例子里，超过这临界值，酒就汩汩地流了。）对化学钟来说，不管X和Y初始浓度是多大，超过临界值，振荡便要出现。反应体很有规则地一会儿变红，一会儿变蓝。这现象，包括啤酒的振荡流出，叫“霍普夫（Hopf）不稳定性”，因为是这位数学家发现的。

这些颜色变化可以用简单术语来描述。这台化学钟可以表示为一个圈或循环，叫“极限环”。我们记住这幅图像就行了：当化学剂从蓝

变红时，它顺着这个环滚动，就像一个轴承滚珠沿着一顶墨西哥宽边帽的边滚动一样（图18）。这个反应可以想象为绕着圈子循环不已，每次经过一个极点，颜色从蓝变红，过另一个极点，又从红变蓝。即使加进去的原料有少许的改变，代表反应的点仍然回到这个有规则的颜色循环，它总会滚入这个环，就像滚珠总是会滚到宽边帽帽边的底圈一样。由于这样的行为，这个环是一个“吸引子”（上章我们已经遇到）。如果停止加原料，让化学反应进行到底，颜色不再变化，紫色液体就重新出现。对这个热力学平衡的情况，吸引子是一个单一的不动点，它可以比喻做反应滚进了一个漏斗的底。在那里，熵达到最大，X和Y都是时间上的常数。

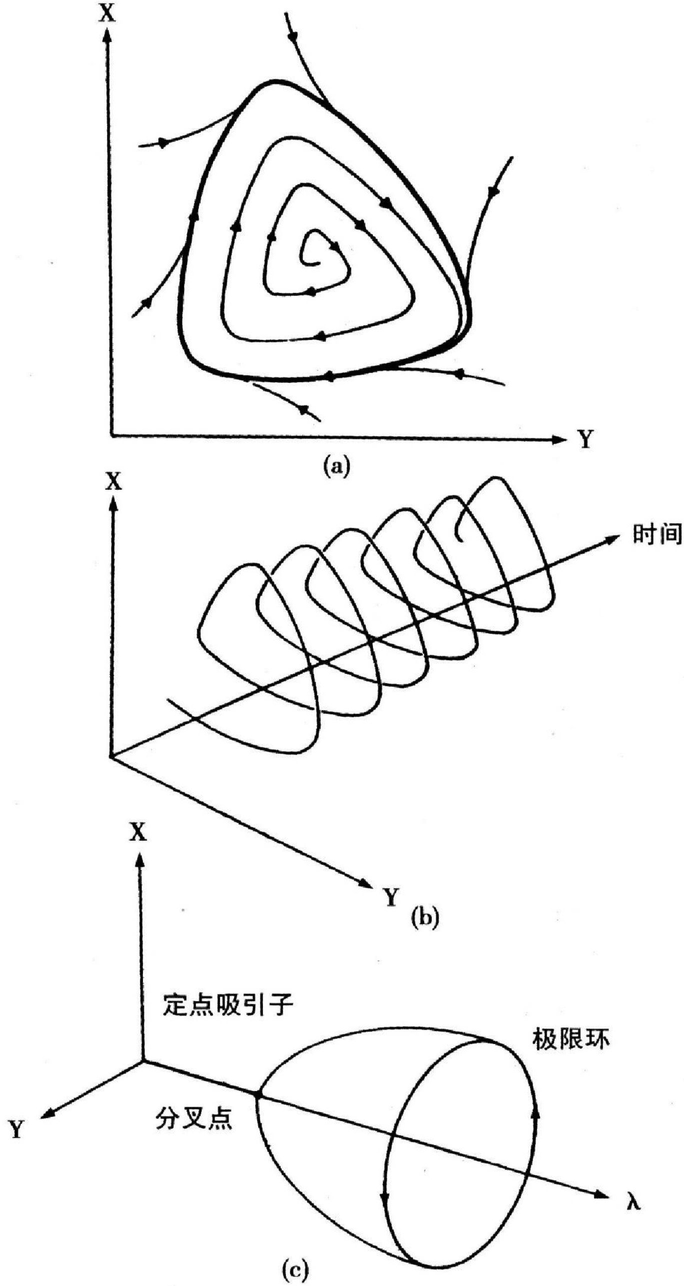



图18　极限环行为。（a）二维投影：X和Y代表化学中介物（好比说，红色和蓝色）的浓度。（b）显示时间上变化的三维图。（c）极限环吸引子从定点吸引子中出现。

现在我们应该好好地回味一下上面讲的。布鲁塞尔振子让我们清楚地看到，如何从无序经过自组织而达到有序。因为把系统保持在远离平衡的状态——办法是不停地加原料，所以反应器中的液体就周期性地由蓝转红，由红转蓝，而不是始终不变的，毫无色彩的紫灰色液体。当然，布鲁塞尔振子只是一个模型而已。然而，它所描述的行为不仅理论上可能，并且我们将会看到，在戏剧性的贝鲁索夫-扎孛廷斯基反应中出现的振荡以及更广泛出现在生物界的各种振荡现象中，它给了我们深刻的启迪。

振荡式的化学反应和一般的化学反应有什么不同？平常的（线性）化学反应可以比为汽车制造，工厂外面停车场上的汽车不断增多，仓库里的零件不断减少。在这过程中，中介物——部分造好的汽车，它们的数量差不多不变。一个试验管里，起作用的原料逐渐被消耗成为产品，分量越来越少。在这原料不断减少，产品不断增加的同时，存在着一定少量的中介物。

在一个非线性的化学钟反应里，反应物的浓度仍然减少，产品的浓度仍然增加，但只要反应物的浓度保持在某个阈值之上，在原料转化成产品，将反应带入平衡态的同时，中介物的浓度（也就是说，混合体的颜色）便会沿着一个极限环，有规则地振荡。这个过程和上面已经遇到过的啤酒汩汩流出，属于同一类。只是在布鲁塞尔振子的反应物保持一定的输入和输出的情况之下，颜色的变化才能持久下去。

这个行为和瑞利-贝纳不稳定性引起的蜂巢结构，同样地令人惊讶。布鲁塞尔振子里面所有的分子都能越过大距离互通信息：它们都知道什么时候变蓝，什么时候变红。沿着极限环滚动时这台钟的“滴答”，只是布鲁塞尔振子某些物理性质的函数。它和初始条件完全无关。这当然是耗散式结构的另一个例子（参见第五章），“耗散式结构”这一术语强调它源于一个和时间之箭有关的不可逆（远离平衡）的热力学过程。

世上常有这样的事：数学家由于数学上的兴趣而不是由于科学上的原因，先就把有关的概念研究好了。法国数学家庞加莱和（前）苏联的安德罗诺夫（Andronov）学派早先就研究过这类耦合微分方程，并取得了关于极限环行为的结果（“极限环”这名词就出现在庞加莱时期）；布鲁塞尔振子使这门抽象数学活跃起来。

耗散式结构的概念在许多领域里受到欢迎。它促进人们从科学角度，而不是纯粹从数学角度，对非线性微分方程发生兴趣。人们研究时钟式反应的化学性质，是因为这种反应容易控制，也比较容易模拟。这方面的努力倒过来又为单细胞和多细胞群体的生物学过程的数学模拟铺开了道路，并且启示我们，同样能用极限环描述的时钟式反应的生物化学“近亲”，会对生命具有重大的意义。生物化学钟像是有机体生命调制过程中的一部分，因而大力推动了非线性现象的研究。

研究非线性系统的数学方法一旦普及，这些方法就涌进了物理学、化学和生物学，导致了一批令人瞩目的跨学科研究。的确，演化性系统的数学描述需要能跟随事物瞬息变化的微分方程。由于反馈是非线性行为的配方中如此重要的一部分——就像化学钟里有些分子，由于自生、自灭或者竞争，而参与自身的命运，类似的分析被应用到“软科学”里，诸如社会生物学、社会学、社会经济、经济学等；在那些领域里，反馈也是存在的。从这些目前热门的分析工作里，又出现了一个时髦语汇——人为生命。按照郎顿（Christopher Langton）的说法，“人为生命”是研究某些人造的系统，这些系统表现着活体系特有的行为；这门研究，不仅对现有的生命，而且对可能的生命，都会给我们一些启示。

## 化学图案与化学波

非线性化学钟模型中隐藏着许多秘密。时间上的图案，我们已经讲过。空间的图案呢？前面分析布鲁塞尔振子从而得出极限环的结果

时，我们忽略了另一个不可逆过程——扩散，和它可能起的作用。那里我们假设了反应器里的化学组成搅得很匀，各种成分A、B、C、D、X和Y，都是很均匀地分布在浑汤里面。如果我们回到汽车制造厂的比喻，那就有点儿像假设汽车如雨后春笋，在工厂到处长出来一样。这显然是不合实际的，比较实际地说，布鲁塞尔振子反应里四面乱跑的各种化学成分是需要时间才彼此相遇的。如果反应器没有搅动，我们肯定不能假设中介物X和Y，产品C和D都自动地、同样多少地在器皿的每个部分形成。因此我们应该设想，反应器里的各种成分先是东一堆，西一堆，它们必须迁移到别处才能参与反馈作用。我们必须考虑在这种情况下，会发生什么现象。

这里牵涉到的是分子的运筹问题。我们要模拟的是：这些起反应的分子在彼此相遇而起作用以前，如何在反应器中运动。答案很简单。只要加进以费克（Adolf　Fjck）命名的定律里的一项，就可以把扩散的影响包括在分析里面。费克定律给出个别成分在空间某点的浓度与其在时间上的变化之间的关系，对汤里每个成分给有一项。这两者之间的转化率叫扩散系数，每个化学成分系数不同，因为块头大的分子扩散得慢，苗条的分子扩散得快，并且还要考虑到它们在其中运动的溶液的黏滞性。

借助费克定律，可以把化学混合体在时间上的行为和在空间出现的图案联系起来。用术语来说，现在模拟系统用的是“偏”（而不是“常”）微分反应-扩散方程。关于这一点，我们不必太关心，仅仅注意一下，这里的数学更加复杂，相应的物理化学内容也就更加丰富。例如，由于霍普夫不稳定性而引起的极限环现在不仅在时间上，并且在空间也可以转动。有一种在时间空间都变的东西是大家都熟悉的——波浪：想一想海滩上的海浪。的确，在霍普夫不稳定性统治之下的化学钟里，我们应该预期反应器里出现的是一条一条红色和蓝色的波纹，而不是整个液体同时变红，同时变蓝。

也有可能演化为一个固定的空间图案，不再随时间变化。这种过

程，数学生物学家用来说明许多现象，例如斑马是如何得到斑纹，蝴蝶如何得到翅膀上的图案，某些化学配料如何使一团一模一样的卵细胞发展成为一个胚胎。在布鲁塞尔振子的情况下，只要先加一些化学剂，再让扩散发生，于是点纹、条纹就会在试管中出现。这就是我们已经遇到过的，下面第七章还要讨论的不稳定性。

经过布鲁塞尔振子的例子，我们看到了各种自组织行为的组合。它告诉我们，非平衡态的布鲁塞尔振子可以在时间、空间、或者同时在时间与空间自动组织起来。这些概念对生物学的意义极为重大。因为在你的身体中，就当你现在念这个句子时，一大批组织在时间和空间的过程就在进行，包括眼睛的运动，心脏的跳动，乃至脑中神经细胞的激发。

目前我们继续把讨论限制在化学方面。我们不应该忘记，我们一直在讲的布鲁塞尔振子只不过是个模型而已，我们采取它是由于数学上的方便。尽管如此，对于布鲁塞尔振子的研究为我们的下一步铺好道路，那下一步是：了解并接受各种式样的自组织现象，这些我们即将考虑的现象当初好像跟热力学第二定律的含义相冲突。

## 诱人的贝鲁索夫-扎孛廷斯基反应

贝鲁索夫-扎孛廷斯基反应的故事就和它的名字一样令人寻味。美国研究该反应的一流权威温弗利（Art Winfree）告诉我们，贝鲁索夫在20世纪50年代的初期，在属于（前）苏联卫生部的一个实验室当头头时，做了该反应的关键工作。在他的研究中，他配制了一种奇怪的化学剂，目的是想模仿克雷布斯（Krebs）循环，从而对它取得某些了解。克雷布斯循环是活细胞把有机食物分解成能量（以名为腺苷三磷酸分子的形式）和二氧化碳的必经之路。

贝鲁索夫模仿该循环的反应含有以下几种配料：枸橼酸，这是克

雷布斯循环的实际成分之一；溴酸钾，其目的是模仿枸橼酸燃烧（氧化）后的生物学后果；硫酸；和一种铈离子的催化剂，因为他觉得这跟许多酵素的作用有几分相似。（化学反应中，酵素的“活动地点”经常带有一个带电的金属原子。）使他惊讶的是：溶液开始在无色和淡黄色两种状态之间变来变去（相当于带电铈离子的两种不同的形式），而且变化是像时钟一样地有规则。贝鲁索夫在他随后的研究中，可能也观察到空间图案的形成。这样，贝鲁索夫首次提供了一个实在的化学反应，支持反应和扩散的双重不可逆过程可以产生自组织的概念，这概念差不多在同时被杜灵从理论上预测到。然而，就如温弗利不久以前所写：“这反应的古怪行为根本是30多年来理论化学家和生物学家从未想到的。”

对贝鲁索夫不幸的是，这个反应是如此奇特，使他很难说服科学界这是真实的。他1951年年底的一篇稿子就被拒绝。编者对他说，他“所谓的发现”是绝不可能的。6年以后，贝鲁索夫又投了一篇更全面的分析，而编者只肯发表一个经过大量删节的短讯。贝鲁索夫的工作最后终于悄悄地出现在一个辐射医学学术讨论会的文集里。论文只有两页，出现在他本人另一篇论文前面。

那时科学界被对第二定律的朴素理解——有序单调地退化为无序，弄得如此昏聩糊涂，以至没有人肯接受贝鲁索夫有关化学系统能自发出现自组织的报道。人们以为第二定律是说任何化学反应总是走向退化的平衡态。而一个来回于两种颜色之间的化学钟意味着反应居然可以走回头路，这不是跟第二定律开玩笑吗？（事实上，贝鲁索夫并不是第一个受这种冤枉的人：加州大学伯克利分校的布赖（William Bray）1921年在过氧化氢转化为水的过程中，也发现了一个振荡式的化学反应，他这发现被认为是由于实验操作低劣而产生的人为现象，而未被接受。）

对于这个反应的兴趣，只是当扎孛廷斯基学习了贝鲁索夫的振荡式配方以后才开始的，尽管开始很慢，而且当时是限制在铁幕后面。

20世纪60年代，扎孛廷斯基以莫斯科大学的生物化学系毕业生的身份，对贝鲁索夫的基本反应做了一串零星的修改，例如用一个含铁的反应剂代替铈离子，使颜色更鲜明地从蓝变红。这样，他渐渐取得保守派同行的欣赏。别人也开始研究这个奇妙的系统；最近20年来，自组织化学反应的研究已经成为很时髦的一门学科。1979年，有人要求世界各地的科学家，为这项工作的重要性出推荐书；1980年，贝鲁索夫和扎孛廷斯基两人，跟克林斯基（Valentin Israelovitch Krinsky）、伊凡尼茨基（Genrik Ivanitsky）、扎伊金（Albert Zaikin）一齐荣获列宁奖。不幸的是，远在国际上对他的启发性工作承认以前，贝鲁索夫在1970年就去世了。

贝鲁索夫的发现，和其后发展出来的各种变例，现在统称为“BZ反应”。BZ反应很容易做，效果也很可靠。（有兴趣的实验者可参阅温弗利的有关论文。）从这个魔术一般的配方可以得到各式各样美丽的现象。本书黑白插图部分有一套图，显示BZ反应各个阶段。对此出色而复杂的反应，许多人做了深入的研究，也写了整本的专著。整个反应牵涉到30多种不同的化学品种，包括一些短寿命的中介物，它们的作用是作为各种连锁循环反应之间的阶石。这些反应被美国俄勒冈大学的一个小组——菲尔德（Richard Field）、柯乐斯（Entdre Koros）、诺耶斯（Richard Noyes），提炼为一个具有11个步骤的化学反应机制，比4个步骤的布鲁塞尔振子复杂多了。仔细检查这11个步骤，就可以找到一个物体影响它本身制造的证据。这证明存在着自催化，而自催化是反馈和非线性的关键成分。从一大批复杂的中介程序里，俄勒冈小组又提炼出一个简单而重要的只有5个不同步骤的模型，科学界同行给它起了诨名叫“俄勒冈振子”。俄勒冈振子这个模型是对BZ反应演化的理论描述，它能在许多方面描述实验者得到的钟表式行为，包括产生化学振荡的极限环吸引子。

如果我们坚持把学问分门别类，我们就得把这迷人的BZ反应划归在“无机化学”一门。为了对该反应了解得更细致，许多化学家更深入

地探讨了无机世界。例如，伽利略高等学校鲁克斯实验室的布利格斯（Thomas Briggs）和饶谢（Warren Rauscher）在过氧化氢、丙二酸、碘酸钾、硫酸锰、过氯酸的混合体中，发现了振荡，颜色在蓝红之间做周期变化。这种振荡式反应发现日益增多，而它们所遵守的一般原则，现在可以说已经完全了解。这种化学钟其他的例子相继出现的有日本京都的“K模型”，美国印第安纳大学的“IU振子”和“泡沫振子”，它所描述的化学反应能产生一串一串的气泡。

BZ反应的一个重要方面是它具有所谓的“可激发性”。这是指在某些刺激素的作用之下，图案就会生长出来，否则介质就完全平静。一些诸如布利格斯-饶谢反应和以二吡啶钌为催化剂的BZ反应的钟表式反应，在光的照射下，便会被激发，开始自组织活动。可激发性，这种能推动BZ反应的性能，是杜灵完全不知道的；即使在今日，还是经常被理论学家所忽略。的确，“可激发性”的定义仍不太清楚。

数学家、物理学家、生物学家仍继续在探索BZ反应中的奥妙；他们这样做不是没有理由的。因为我们很快就要看到，我们不可能忽略BZ反应和有机世界中许多我们熟悉的组织之间的关系。化学钟里形成的螺旋波与心脏病发作时的波动、原始黏菌（见黑白插图）、旋涡星系、飓风等之间大有相似之处。温弗利甚至写道：“虽然‘BZ反应’谈不上具有一个可以变异，可以演化的遗传系统，它有不少特点，就是使我们对生物体系感到兴趣的特点：诸如化学的新陈代谢（有机酸氧化为二氧化碳），自我组织的结构，有节奏的活动，在某些极限以内的动态稳定，在这些极限以外的不可逆的解体，一个自然的寿命，等等。”这样，关于化学钟的研究可以说的确把无机化学搞“活”了；在这以前，这门学问往往太缺少理解，太多集邮式的、大量资料的盲目搜集。

自组织的无机系统涉及众多的简单化学品种。但其中的化学情况具有较大的偶然性——既然所有在汤里兜圈子的分子都多少可以相互起作用，就很少有特定性。我们将要看到，有机体的可能性倾向于另

一端。在那里的（生物）化学既复杂，又是细致调节的：每个反应都是非常特殊，都是以惊人的效率进行。普里高津和司坦厄斯（Stengers）评论道："这很难是偶然的。这里我们遇到的是区别物理和生物学的一个基本性质。生物学系统是'具有过去'的。它们的组成分子是某种演化的结果；这些分子被选来参与白催化机制，从而产生具体的自组织方式。"这是具有目的的化学。这是生命的奇迹。

## 漫谈分维、奇异吸引子、混沌

在一个化学钟里，非线性的复杂性显示为时间上有规则的行为：起化学反应的混合体的颜色有节奏地变来变去。上面已经看到，描述这种行为的是一个极限环式的吸引子，化学反应在那里的行为，像一个轴承滚珠在一顶墨西哥宽边帽的边缘上滚动一样。我们应该把这种行为和描述热力学平衡的定点吸引子对比[图19（a）]；定点吸引子我们前面曾比作一个漏斗的底。

然而由于不可逆过程而产生的还有另外一种吸引子，它描述的是时间上完全两样的行为——混沌。退化为混沌的过程最好用分叉图[图16（b）]来说明。分叉图显示当诸如化学钟的一个系统被推得离平衡态很远以后，它各种可能表现的行为：在第一个临界点，它分枝为二，产生两种可能性。每一枝又依次生出多个小枝，这样枝上生枝，一直生下去，数学上这相当于非线性系统能在完全一样的情况之下表现多种不同的行为。这些临界点或分叉点越来越多，最后整个图的右方便是许许多多的可能性密集的一团。让我们回想一下横坐标的意义："图的右方"就是代表远离热力学平衡的地方。

由于巨大数目的可能状态紧紧地聚在一起，在此场合，可选择的行为之多，令人眼花缭乱。系统已不再只是限制在少数几根"枝"上，而是可以在无数的可能状态中取样。一个系统要从平衡态（横轴的原点）到达这样的混沌状态，在它被推向离平衡态越来越远（但不是无

穷远）的过程中，它可能经历了无穷多个临界点。我们或许会以为，离平衡态越远，这棵分叉树上的混沌便越普遍。然而复杂的程度远大于此，因为分叉树很像法国梧桐，每层树叶之间仍是空的。这样，在混沌里面存在规律性的“岛”或“窗”，窗里又有窗，一直下去，无穷无尽，并且“反之亦然”。本章下面还要重游个别通向混沌的道路。

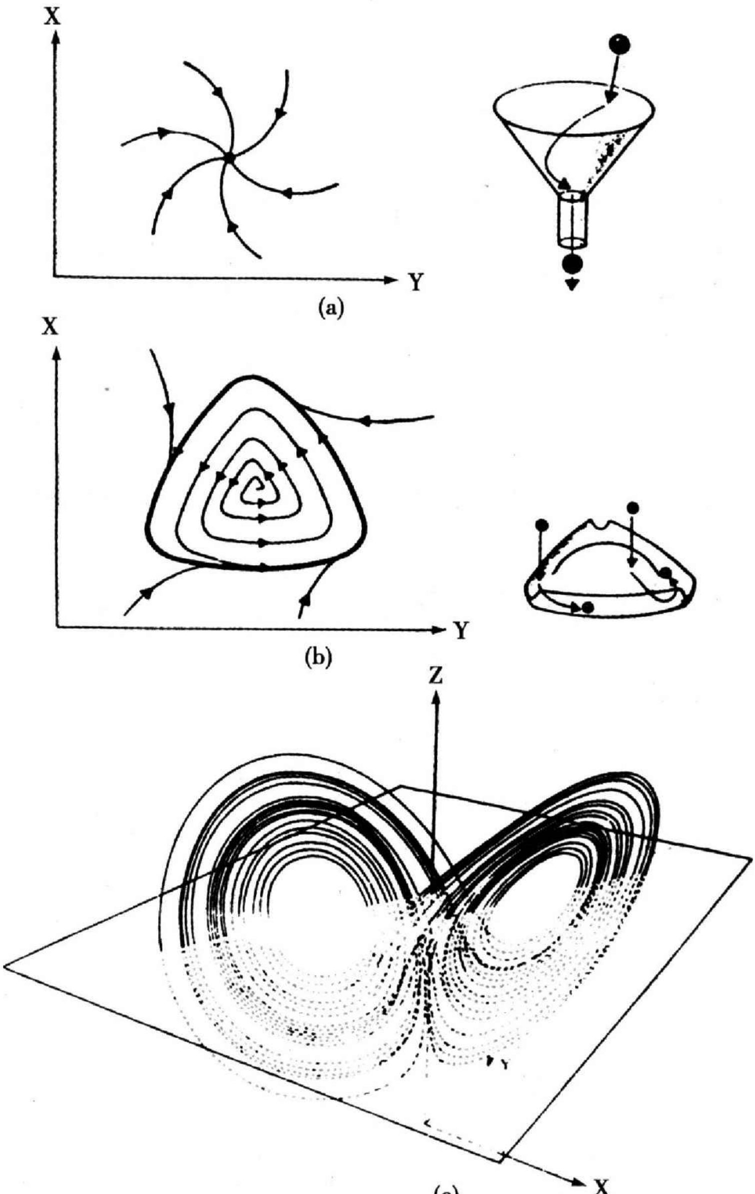



图19　三种吸引子。（a）定点吸引子（稳定；平衡）与其机械对应体。（b）极限环（周期性）吸引子与其机械对应体。（c）洛伦兹奇异吸引子。

混沌演化看上去和我们一直在讨论的完全相反：它否定时间演化中任何长期规则性或可预言性。一个化学钟的成分浓度如果有了改变，如果它被推得离平衡态太远，它的颜色便不再出现一次一次有规则的变化：它变成一个混沌混合体了。在这种情况下，它变红变蓝完全是随机性的：我们不能预言下一次变化是什么时候发生。某一次的实验记录结果不会重复。下一次实验会出现另一套随机的颜色变化的时间间隔。

尽管有这种不守规矩的行为，混沌还是可以用吸引子的概念来理解。这一点是茹厄勒（David Ruelle）和拓肯斯（Floris Takens）在1971年证明的。茹厄勒出生在比利时，在巴黎附近的伊菲特河上布若镇的高级科学研究学院工作；拓肯斯则来自荷兰赫罗宁艮大学。他们的论文题目是："关于湍流的本质。"论文提要短得惊人——"本文提出耗散式系统中产生湍流及其有关现象的一个机制"；而论文本身却是密密层层的高级数学。两位作者想理解的是，例如当你把水龙头大大打开时，初始平滑的流动如何转变成本质复杂的湍流。可是他们结论的应用范围远远超过这些例子，出现的是一个怪兽，叫"奇异吸引子"。

这跟实际世界有何关系？茹厄勒用香烟的烟在宁静空气中的上升的例子来说明它的用途："烟柱在一定的高度上出现振荡，振荡是如此复杂，要理解它看上去几乎不可能。虽然它在时间上的演化遵守严格决定性的规律，它的行动却好像是自己做主。物理学家、化学家、生物学家，一如数学家一直在想了解这种情况。在此过程中，他们从奇异吸引子的概念和现代计算机的运用里，得到帮助。"

奇异吸引子的来源。茹厄勒描写如下："我问拓肯斯，这个极为成功的词语是不是他创造的。他回答说：'你问过上帝是他创造了这该死的宇宙吗？……我什么也记不得……我常常创造，过后就不记得了。'这样看来，奇异吸引子似乎是在狂风闪电之下诞生的。不管怎样，这个名字很美，极适合那些令人惊讶而我们还很不明白的东

西。”另一方面，英国数学家塞曼（Christopher Zeeman）认为：“或许一个更好的名字是‘混沌吸引子’，因为现在它们中间许多例子都不太奇异了。”这两个名字都有人在用。

奇异吸引子和我们先前遇到的两种吸引子——定点和极限环，大不相同（见彩色插图），虽然它也是稳定的，也是代表某种系统可能驻留的状态，也是时间之箭可能的目标。它有两个特性。一是和极限环不一样，它对初始条件极端敏感：一个被一个奇异吸引子捕获的系统，它的长期行为和它当初最细微的细节都有关。奇异吸引子和极限环不同的第二点是：奇异吸引子是一个“分维体”。

“分维”这个词是1975年问世的。曼德布罗特（Benoit Mandelbrot）创造了该词，为的是要描述在不同尺度上都具有同样的不规则形状的奇怪几何。奇异吸引子，不管我们把它的某一部分放大多少倍，它基本上仍具有该吸引子的全盘结构。花纹里面有花纹，那里面又有花纹，一直下去，永无止境，这个性质叫“自相似”。同一个花纹在每个尺度上都存在：一片枫叶的边缘上满布着小枫叶形状，小枫叶的边缘上又是更小的枫叶形状（参见黑白插图）。这叫做尺度转换下的不变性，因为物体不管在哪个尺度上看，花样的形式都是一样的。

曼德布罗特的工作撼动了我们对维度和维数的想法。众所周知，线的维数是一，而正方形里的面是个二维体。但是实际上，这些差不多总是理想化过的：物体的维数可以是一点几。此处的“点几”就是说该物体的维数是一个分数。曼德布罗特为了说明这个观念，在他的一篇论文里问：“英国的海岸线多长？”稍思片刻，我们就知道答案跟用来量海岸线的尺度有关。用海边城市之间的直线距离，我们算得的是一种粗略的估值。但你如果沿着海岸步行，绕着每个小海湾、每条小河的出口走，你就会发现这海岸线大大地增长了。对一个蚂蚁来说，仅仅小石头就要大大地拉长旅程，至于对一个蠕动的细菌，英国的海岸简直是永无止境。答案很明显地和测量所用的尺度有关，这是因为

基本上在所有的尺度上都存在有结构。的确，如果我们能把尺度缩到无穷小，海岸线的长度就会变为无穷大。因此我们有如下的似非而是的结果：海岸是一条无穷长度的“线”，很容易地包含在一个有限的面积里面（围英国画一个圆）。

实际的海岸线具有自相似的分维性质，虽然这句话应该从平均、统计的角度去理解。有一个用数学定义的曲线和海岸线十分相似，叫“苛赫曲线”[1904年苛赫（Helge von Koch）引入]，它由一系列越来越小的三角形组成，如图20所示。苛赫曲线的维度介于一维的欧氏线和二维的平面之间，它的维数的近似值是1.2818。

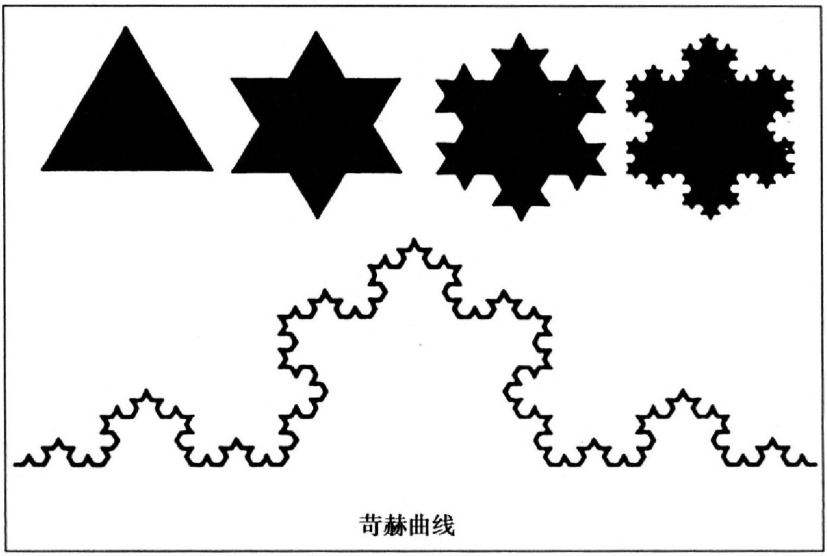



图20　苛赫曲线。作法：开始时三角形。在它每条边上加一个新的小三角形。这样继续下去做成下方的曲线。

分维图案的发现，揭示了一条认识自然界美妙而无穷尽的复杂层次的新途径。曼德布罗特的工作，一如他以前一些数学家的工作，很适合描写我们周围和我们体内的各种自然形态。云和海岸线都是分维体。并且分维体并不仅限于无生物。一棵树根系的二维投影，一幅神

经照片和路过的卫星拍摄的河三角洲的图像都极为相似。它们都可以被认为是分维体，它们彼此相似，是因为它们的大尺度形态可以从不断重复一个简单的数学规律而生长出来。我们身体里许多结构都是由分维组织所控制。曼德布罗特写道：“肌肉组织……不管多小，都具有交叉排列的动脉和静脉。它是一个分维面。”至于人脑的皱褶轮廓，曼德布罗特说：“要定量分析这种轮廓，传统几何是无能为力的，而分维几何却是得心应手。”的确，一个有趣的问题是推测自然界仍保持分维性的最小尺度——这可能表明，追求物质的“最终单元”是徒劳无益的。

奇异吸引子跟时间有什么关系呢？部分回答是说奇异吸引子描述混沌演化，而混沌演化，我们将在第八章看到，完全推翻了时间对称的决定性论。第一点要掌握的是：一个化学反应在奇异吸引子中的代表点，由于吸引子的分维性质，将经历一串无穷系列的点（参见彩色插图）。定点吸引子和极限环吸引子的维数分别是零、一、二、三、……等等，而奇异吸引子可以定义为维数是分数的吸引子。茹厄勒写道：“那些团团的曲线，那片像云的点子，一会儿像焰火，一会儿像星系，一会儿又像奇怪令人不安的植物蔓延。这是一个形态等待探讨，妙音等待发现的国度。”奇异吸引子的维数是分数，这事实使我们对它第二个性质——混沌，有了心理准备。奇异吸引子拥有无穷多的可能性，而这些无穷多的可能性全包含在一个有限的区域里：随着时间的流逝，系统取样于不同的位形，永不重复。我们可以想象系统无止境地在描出图案中的图案里面的图案。这乍看上去似乎很难想象。然而一旦有了分维体的概念，就不难看出，一个系统——奇异吸引子，不因为它是限制在一个有限区域之内，就不能跟永无止境的新机会相遇。

一个动力学系统一旦被吸入一个奇异吸引子，该系统的长期未来行为就变为完全不可预测的了。这是因为，如上所述，奇异吸引子对初始条件敏感到难以置信：除非系统以严格的无限高精度开始，它终

究将会变为完全不可预测。虽然控制不可逆系统时间演化的微分方程是决定性的，虽然原则上初始条件一知道就可以预言整个未来，可是系统对初始条件的极端敏感彻底粉碎了可预言的钟表式宇宙的想法。

为了突出这种异常行为，我们可以将它和陷入极限环的化学钟对比。代表化学钟的滚珠不管是怎样扔进那顶高边帽，它最后总是绕着帽边儿滚动。但是在一个混沌奇异吸引子的范围里，发生的完全是另一回事。假设滚珠滚进一个奇异吸引子里面了，而你想要它重复它经历过的那条复杂的路线。你将发现，不管你取哪个邻近的出发点——不管多近，总跟当初的不同，你的轨道会很快地和原轨道分散，在吸引子里面做完全不同的运动，走的是分维体无穷花样套花样里面的另一条轨道。

耗散式混沌产生于奇异吸引子的套中有套层出不穷的世界。对这种混沌系统的实验，只有在以无穷高的精度得知初始条件的情况之下，才有可能做出绝对准确的预言。但实际上不会有这种情况，初始条件多少总有一点不确定性，这不确定性将随时间以指数方式增大。混沌和对初始条件敏感性之间这个关联，极为重要，它使我们可以对时间之箭给出一个自洽的科学描述。

然而，在决定性混沌里面——叫“决定性混沌”是因为它来自决定性的非线性方程——也存在有某些规则性。这种混沌是系统内产生的，是系统的一个内禀性质。因此，在概念上，它和外界环境随机涨落（噪音）的影响迥然不同。这种随机过程——噪音，能在一个并未陷入一个奇异吸引子的系统里面，产生随机的像是混沌的行为。科学家面临的跨栏之一就是，如何区别决定性混沌和随机性混沌。下章我们谈一些复杂的生物现象时，这座障碍又会来挡路。

决定性混沌使“有序”、“无序”的概念变模糊了。近来有一种倾向，用“混沌”（意即决定性混沌）一词来解释一切，不仅用于不可预测的或不稳定的场合，并且用在用“自组织”更为恰当的地方。我们不

要被“混沌”这个时髦字眼弄得眼花缭乱。秩序和决定性混沌来源一样，它们都是用非线性微分方程描述的耗散式动力系统。不过，就如下章所述，对生物学和生命本身来说，有序的情况往往比混沌的情况更为重要。当研究者打着混沌的时髦旗号把论点放在我们面前时，我们应该多少带点儿怀疑态度。对每种情况应该分别加以评价。

化学混沌，茹厄勒早在1973年就首次提出。在我们的化学钟例子里，当颜色从红到蓝的变化不再像钟表那样地有规则时，那便是奇异吸引子存在的标志。茹厄勒告诉我们，为什么决定性混沌被经典科学认为是违背正道。这是因为传统上，科研者从数据中找到规则模式以后，他们就很有希望理解这些规则模式。1971年，茹厄勒问一位研究振荡反应的专家，问他是否碰见过，对时间的倚赖是混沌式的反应。他回答说，以前要是一个化学实验者得到一串混沌式的记录，他就肯定把记录扔掉，说实验没有成功。现在情况总算好些了，现在我们有多个非周期性化学反应的实例了。

混沌能以多种方式在化学中产生。一个配方是：先按照通常办法把振荡式反应建立成一个开放系统，用搅动式反应器使系统保持在远离平衡的状态。这时原料的输入率如果固定，反应便会成为一个稳定的颜色周期循环。现在假设我们提高原料的输入率，以不同于化学钟的频率，改变原料浓度在时间上的变化。我们可以把化学配料的流率作为离平衡态距离的标志：流率越小，反应就越靠近平衡态，流率越大就离平衡态越远。因此，当流率增加时，反应就被推过一个又一个的临界点：大到一定程度以后，混沌式化学便会崭露头角（参见黑白插图）。

位于奥斯汀的得克萨斯大学的斯温尼（Harry　Swinney）与其合作者详细研究了BZ反应的动力学性质，他们得到有力的证据，证明该反应混沌状态中存在着一个奇异吸引子。决定性混沌虽然是化学本身的某种学术性奇物，但对它进一步的了解对化学工程将会有用，因为许多化学工业过程本质都是非平衡的。在有生命的系统中，混沌所扮

演的角色也可能重要，——有人甚至认为是不可缺少的。

奇异吸引子的概念虽然是在1971年才被茹厄勒、拓肯斯明文写出，却早已隐含在麻省理工学院气象教授洛伦兹（Edward Lorenz）1963年的一篇论文之中。洛伦兹想了解天气预报为什么常常不准。英国的一个天气预报员菲什（Mi-chael Fish）肯定会觉得洛伦兹的话很入耳。1987年10月15日，菲什对电视观众说："一位女士刚来电话说，她听说暴风雨就要到了；观众们，请放心，没那么回事。"但果然就有了那么回事。

洛伦兹的工作为这类错误的预报提供了一个有力的辩解。凭着一台计算机（当时还是很稀罕的东西）和他那一行少见的数学本领，洛伦兹致力于设计一个大气气流的数学模型，要它尽可能地简单，但不漏掉任何重要的物理性质。洛伦兹的方程对一层从下方加热的水平液体给出一个近似描述。液体较热的部分比较轻，要向上浮，从而搅起对流。如果加热够强，流动将是不规则的湍流。洛伦兹最后得到的是三个相互耦合的非线性微分方程——要奇异吸引子出现至少要有三个方程。洛伦兹研究了这组方程，逐渐意识到，求解时输入计算机的初始天气条件不管有多么微小的变化，结果（天气预报）就会在很短期间完全改变。要是别人就很可能说这是计算机有什么毛病，但是洛伦兹在气象学上的经验使他能完全接受这个出人意料的结果——在这一点，他是远站在他时代的前面。他的奇异吸引子（现在以他命名）直到十多年以后才得到公认。不过就是今日，这还没有被证明为数学意义中的奇异吸引子，虽然它所有的物理性质都和我们所期望的一样。

用越来越精巧的计算机来取得越来越准确的天气预报，这个想法由于混沌的存在，面临一个严重的障碍——洛伦兹方程对初始条件的极端敏感性，即洛伦兹所谓的"蝴蝶效应"。这生动地说明，由于混沌，最微小的事件会引起最巨大的后果。奇异吸引子，差之毫厘，谬以千里：亚马孙森林里一只蝴蝶抖一下翅膀，就会引起西印度群岛一场狂风暴雨，等等。然而，夸张的比喻说说固然无所谓，但不要忘

记，如果为了更符合实际，我们在洛伦兹方程里多加一些变量，混沌就反而更难找到，而不是更容易找到。

## 级连通向混沌

什么时候可以看到混沌？这个问题相当重要，因为混沌可能是好消息，也可能是坏消息，——看我们讲的是癫痫还是心脏病发作（见第七章）。可是，这问题的全部答案，超出现今我们对不可逆非线性系统复杂无比的行为的知识范围以外。一套包括所有混沌可以出现的场合的理论，仍然是一项巨大的工程。许多科研人员只是满足于在一些模型问题里找混沌，煞有介事地计算所得到的奇异吸引子的分维数（其实这主要的结果只是使科研文献膨胀）。对现今“强调非线性混沌、轻视其姊妹课题——自组织”的态度恼火的人讥诮说，这种系统不管你是研究哪门学科的，迟早总会碰到混沌；但这本身是否有意义，则大有问题。

对于肯定会产生混沌的一些场合，我们有了一定的认识。周期性极限环控制的规则振动状态也好，定点吸引子控制的恒定态也好，它们的破坏都会引起奇异吸引子的产生。前者对生理学有重大的意义：下面我们将看到，当一个极限环的调节作用被破坏而混沌出现时，生物学不正常现象就会相应而起。

上面我们已经叙述过茹厄勒-拓肯斯通向混沌的路线：那里需要系统被驱赶过三个或三个以上的极限环分叉点；这条路一般叫“类周期路线”。从一个极限环的遗迹中，混沌奇异吸引子还能以另外两种方式出现。这两种方式的名字听起来同样神秘：一个叫“亚谐波级连”，另一个叫“间歇性”。对它们要详细描述都相当专门；后者是法国科研者泊摩（YvesPome au）和曼讷菲尔（P.Manneville）在1980年发现的，本书将不论述。

为了说明在化学钟里通向混沌的亚谐波级连（又叫费根包牟级连，费根包牟（Mitchell Feigenbaum）工作于洛克菲勒大学），我们前面已经说过，最好的办法就是用一棵简单的分叉“树”。它显示有哪些可能状态，并且显示当系统从靠近树干的只有少许可能状态的区域，被赶到高高在树顶的混沌的模糊一片时，会发生什么。亚谐波级连式路线和茹厄勒-拓肯斯路线虽然都用分叉图来表达，它们的数学细节和物理细节却很不一样。再者，在亚谐波路线里，化学配料浓度的变化是在同一个循环周期中发生的。假设我们有个化学钟，它以T秒的周期做规则振荡。这时候我们刚过图21中的第一个分枝点，那里树干一分为二。现在假设配料浓度以二倍的速度开始变：实际上，现在钟是被一个周期为T/2的外“力”所驱动了。

让我们继续向上爬这棵分叉树。我们越加快配料的流动，钟离平衡态就越远，我们在图21上就越向右移。超过某个阈值，某个临界点，钟的第一个振荡就变为不稳定，周期就突然转换为（大致）双倍长的新周期。这个新行为，其中颜色变化每周期增大了两倍，图上由第一临界点过后的两对线代表。流率一再的增大把钟依次推过一个又一个的临界点，每分叉一次，周期就乘二，变成4T、8T、16T，一直下去。此过程叫“周期加倍”，是最经常走的通往混沌之路。最后，在某个有限的流率之下，由于无穷多的串联分叉，钟整个解体，达到的是无周期状态，周期无穷大，系统永不自我重复。这时，系统陷入一个奇异吸引子，那里它再也不会重复已经走过一次的道路。这个周期不断加倍的极限和混沌是同义词。这好像是：当可取的时间组织方式太多时，混沌就抛头露面了。

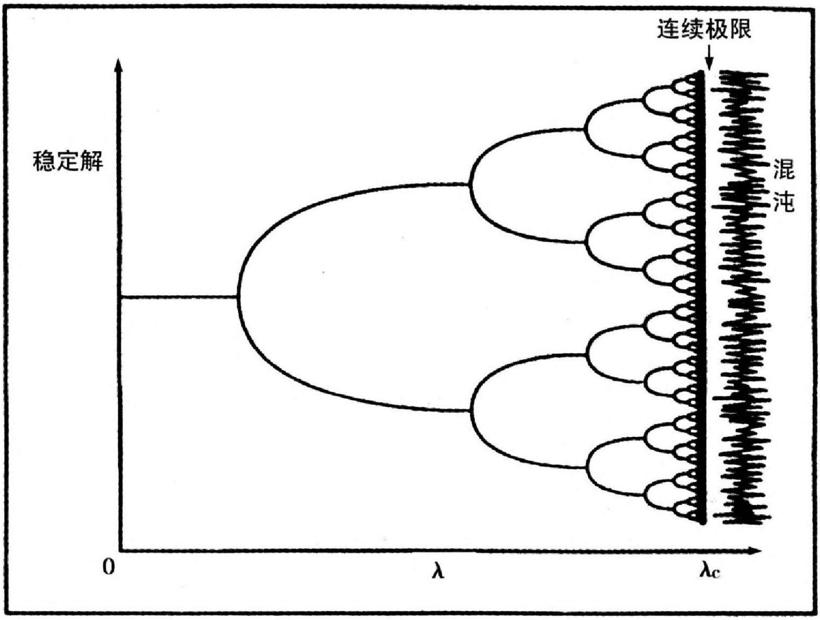



图21　一个简单的非线性系统的通向混沌的分叉级联（周期加倍）。注意分叉的规则重复：在离原点（平衡态）有限距离$\lambda_c$以内，就已出现无穷的分枝。

对这种周期加倍现象的数学性质的理解，许多研究者做了重要的贡献，尤其是梅尔堡（P.Myrberg）、沙尔可夫斯基（A.N.Sharkovsky）、麦（Robert May）、奥斯特（George Os-ter）和费根包牟。对我们来说，此级连最出色的特点就是它的一般性。意思就是：从周期不断加倍而产生的混沌，不管是产生在哪个系统（有机世界也好，无机世界也好，都有许多这样的系统），都具有类似的数字比例关系。该一般性在实验上极为重要：借助于它，我们可以从乍看上去是纯粹噪音的数据里，把决定性混沌清理出来——决定性混沌其实是一种潜在的秩序。

许多研究者认为，有众多不正常的生理状况，对它们的诊断少不了对混沌的正确了解。1980年，茹厄勒关于混沌和心脏的跳动，推测如下："这对我们每个人说来都是关系重大的。正常的心脏状态是周期性的，而许多非周期性的病态会导致死亡的定态。可以复制各种心

脏动力状态的合乎实际的数学模型，对它们在计算机上研究，看上去会对医学上有很大的好处。”下章我们将看到茹厄勒的预感正确到什么程度。

非线性系统中的分叉点或临界点有一个特性，很清楚地说明哲学家伯格森（Henri Bergson）在他著作中大力鼓吹的论点——时间是“创新的介质”。一个系统在其分叉图上的位置同时反映它个别的历史：就像每个小孩儿都知道，要摘树上一个苹果必须爬过一定的树干，一定的大、小树枝；因此，如果系统在分叉图里没有选择某个特定的路线，它就不会到达它目前所在的地方。因为决定临界点结局时，不确定性和随机性扮演了主要角色，所以时间便成为一个创新的实体：从一个稳定状态到下一个稳定状态之间，系统整个的未来都悬于机遇，这和系统的过去是不同的。分叉图所揭示的时间不对称性和我们所体验到的一样：一个一星期大的婴儿会长成为一个王子或者一个叫花子，但一个五十岁人的历史是固定的。同样地，设想一个甲虫在分叉树上爬上爬下。它可以随便从哪个树叶爬到树干。但要从树干爬到某片树叶，它必须在树枝中取一个特定的路线。这样，甚至停在分叉树上小树枝的甲虫都具有一个特别的历史。

## 对有机世界的一个物理化学观点

现在我们到达了一个重要的转折点。本章和上章讨论的诸如化学反应等过程的时间演化，总归是以热力学第二定律为基础的。由于包含在第二定律里面的时间之箭，我们看到非平衡过程在“无机”物质中既可以产生自组织，又可以产生（决定性）混沌。

在第五章我们看到，有生命的动植物是存在于远离平衡态的条件之下。让我们来考虑一下神经脉冲是怎么形成的——读者念这句话时就需要千千万万同步运行的这种脉冲。归根到底，一个神经细胞的激发，有赖于在细胞膜的两边，借助于一系列的分子泵，把钾离子和

钠离子隔开，使钾离子集中在膜的里面，钠离子集中在外面。细胞膜里有专门的离子通道，它们自动弹开，让离子沿着浓度梯度流动；通道的交通规则是使离子流以脉冲方式流进细胞，就像啤酒汩汩地流出酒瓶一样。很显然，这是一个非平衡情况：如果是平衡状态，那在膜的两边都是两种离子的均匀混合体了。我们现在可以理解，为什么平衡态就等于死亡——浓度梯度一旦不存在，就无法激发神经细胞了，就不会有思想了。

很自然地，我们要把生命的无比复杂性看做是自组织过程的结果。有了自组织的概念，就不难想象时间中、空间中的有序结构以及在某些适当情况下的决定性混沌。令人注目的生物学秩序可以在这种自组织的基础上取得了解，这样的实例现在已经很多；生命的基本过程，可以用承认时间是不可逆的非线性微分方程来解释。

这些生物科学中的进展有一个特点，很令人寻味——那就是，这些进展来自研究非线性问题的数学家和物理学家之间的经验交流。直到20世纪60年代后期以前，存在有严重的障碍：“语言隔阂”，行话的浓雾使某一领域的工作者看不懂另一领域的成果。此外还要加上另一个双重因素。一方面，科学家不喜欢生物系统令人头昏脑胀的复杂性，它使设计可以检验这些系统的理论和实验的工作，更为困难；另一方面，生物学家对数学推论和传统怀有戒心，认为数学模型总是把问题过度简化了。但为了抓住控制生物系统中自组织的基本原理，今日大多数生物学家都承认“漫画式描述”的必要。我们如果要做的话，当然可以试把所有各个细节都包括在我们的现实模型里面，但这是非常艰巨的工作，并且无论如何，十之八九是会把事物的主要真相弄得更不清楚。

## 分子演化，复制和生命的起源

经过若干年辛苦的实物研究，达尔文得到如下结论：现今所有的

物种在几十亿年以前，都有同一个祖先。这个所有生物为其后裔的老祖宗，一定是一个由单细胞或少数几个细胞组成的有机体。达尔文的基于变异和竞争选择的进化论，经过生物学家不断搜集资料，日益巩固了。但是，那最简单的生物又来自何处？达尔文没有答复这个问题。有人提议过上帝，然而现代科学对神的干预的观念，是不太客气的。

难道说没有一个自然过程，能使一个单细胞从无机体产生出来吗？早期该观点持有者之一是耶稣会教士泰雅德沙丁（Pierre Teilhard de Chardin 1881～1955），他一生的目的是想把科学和宗教融合为一。他认为有机体和无机体都随着时间的流逝，逐渐组织成越来越复杂的形式。这应该是关于自组织最早的想法之一。不幸的是，为了他的观点，泰雅德沙丁付出了很大的代价。1924年，耶稣会禁止他在巴黎天主教研究所讲课。1926年，他离开法国去中国流浪，最后死在纽约。

关于这种演化可能在何处发生，最有影响的早期想法之一出自达尔文本人。在他1871年写的一封信中，他写道：“但假设（当然这是个很大的“假设”）我们可以想象，在一个具有各式各样的氨盐、各式各样的磷酸盐、日光、热、电等的温暖小池塘里，化学反应形成了一个蛋白化合物，接着又起更复杂的变化，……”

从现今观点来说，我们原则上可以理解，物质在时间上和空间里的组织——有机体显著的特征，经过远离平衡的不可逆过程，是可以出现的。每个细胞都是一个组织良好的工厂，在里面，惊人的化学反应在化学配料某种极不均匀的分布之下进行。在此层次可以看到的美，对神经系统做过前驱工作的卡哈尔（Santiago Ramony Cajal 1852～1934）撰文描述如下，其中生物学术语较多，提及的知识也稍嫌过时，然而文笔确是华美。此文出现于1937年出版的卡哈尔自传——实验生物学中的一部经典著作。这里，卡哈尔描写他在显微镜里观测到的世界：

气管田里、喉咙田里种满着颤动的纤毛，纤毛由于隐藏着的刺激而波动，好像寒风吹进麦田；精虫不倦的鞭泳，气也来不及喘地奔向它情之所钟的卵子；神经细胞，最高等的有机元件，像章鱼一样伸出巨手长臂，一直伸到紧邻外界的边区，提防物理化学力不住的偷袭；建筑简单而严峻的卵子，看守着有机形态的秘密，它的星云状原形质中，围绕着胚胎旋转的是无数个世界，在未来周期里将要出现；肌肉纤维，一种高度复杂的动电电池，在它齐整的结构里，就像在机车里一样，热能转化为机械能；腺细胞简单地为活化学厂制造酵素，为了兄弟元素的利益，消耗自身的体物；脂肪细胞，家庭经济的模范，为了预防未来的饥荒，把生命宴席剩余的食品储存起来，以防备他日某器官罢工或营养发生危机。此等现象，如此多彩，如此协调，强烈吸引着我们，对它们的默想，使我们的精神充溢着最纯洁、最崇高的满足情绪。

读过这段对微观世界一气呵成的描述，谁还会怀疑，生命存在于远离平衡的场合，那里到处都是变化？卡哈尔所描述的结构是可以出现的，条件就是46亿年前地球形成以后的早期，存在有恰当的自组织配方。我们要问：发生了什么事使一个不毛的、无生命的地球变成我们现在见到的样子？完全肯定地，我们知道很少：但是下面讲的，虽然有些地方是推测，还是多少有些道理。

那时，地球的大气由氢、氮、二氧化碳、甲烷、氨、硫化氢和水组成，但很缺氧。一般的想法和其他关于化学演化的看法一致，是考虑这些简单的分子如何可以整理成较复杂的分子。“生命前合成”的经典实验之一，由尤雷（Harold Urey）的一个学生米勒（Stanley Miller）1953年在芝加哥大学报道。米勒把他认为与原始大气类似的东西混成汤，放在一个缸里通电模拟闪电，发现缸里形成了某些氨基酸；我们所知道的生命少不了蛋白，而氨基酸是蛋白的基本元件。从那时起，已有大批的证据，说明一整套的生物学上重要的分子，包括基本遗传单元、核酸、酵索和诸如腺苷三磷酸的储能生物分子，都可

以用类似的方法制成。这些证据主要是马里兰大学的彭恁帕如摩（Cyril　Ponnemparuma），位于圣地亚哥的萨尔克研究所的奥尔格（Leslie Orgel），迈阿密大学的福克斯（Sidney Fox）等人取得的。

不足为怪，关于这些有机分子在无生命情况下的形成，别人也提出过另一些理论。其中之一说，这些简单分子先形成于太空，在叫做暗星云的气体和尘埃组成的云里，然后经由流星带来地球，把这些分子丢放在某些像泥塘这类有利的场所。有人进一步建议，说黏土不仅是制造这些简单元件的催化剂，并且本身就是早期生命形式的一部分，就是由遗传物体DNA（脱氧核糖核酸）和RNA（核糖核酸）控制的今日生命的铺路者。然而不管形成这些初始的复杂分子走的是哪条路，简单分子如何聚合成细胞总还是个问题。这过程牵涉到至少三个因素的演化：一、必须有一层膜把细胞本身和外界分开；二、必须有一个同化作用，由一套协调的（生物）化学反应组成；三、必须有基因，来指挥这首交响曲。

传统的看法是：开始是这些相互作用的分子被关在个别的结构里面，这些结构彼此之间的边界是半渗透性的，这样便允许复杂分子在时间上和在空间里演化了。有奥帕林（Alexan-der　Oparin）的“凝聚模型”，那里水滴围绕着带电粒子形成；有福克斯提出的过程，可以使氨基酸自我组织成微观小球；还有哥岱科（Richard　Goldacre）的“类脂双层体”模型，那里脂肪分子联合力量，制造简单的膜状结构。目前看法强调RNA高分子的自催化功能是细胞膜形成以前的第一推动者。此种功能的发现使得耶鲁大学的阿尔特曼（Sidney Altman）和科罗拉多大学的捷克（Thomas　Cech）荣获1989年的诺贝尔化学奖；在他们的工作以前，所有生物学催化剂都被认为是蛋白。

如果我们从自组织原则出发，我们就可以对可能发生的情况，采取另一种和上述看法互补的看法。如果在生命出现以前的原始浑汤里存在有某种恰当的反馈机制，实现自组织的一般条件便成熟了。例

如，如果浑汤里某种分子能催化自身的产生，非线性反馈——自组织的标志，便出现了。从而介质的均匀性将被破坏，引发出图案和节奏（可能经过类似杜灵1952年提出的途径），就像化学钟能显示时间上和空间中的图案一样。我们由此得到启示：应该力求一种机制，能耦合扩散和适当的非线性（生物）化学反应。

原始浑汤里的某个关键成分于是变为催化自我产生的一种或多种分子：自催化提供了非线性所需要的正反馈，虽然也有别种可能，例如更复杂的“互催化”，其中反馈是经由一系列连锁反应间接提供的。原始浑汤的性质究竟如何，仍在激烈争论之中，好在这性质是远在直接观测的范围之外。不过这里重要的只是原则性问题。对此，奥尔格和他加州萨尔克研究所的合作者做了极有意义的实验。他们证明了，上面提到的核酸具有自复制这最重要的性质：在核酸原料的纯粹化学混合体里，会有更多的核酸形成。

## 遗传钟

核酸掌握着生命的设计。在DNA和RNA里面的是基因，它们逐字给出具体的指令，为我们地球上的生命建造蛋白。这种化学的信息技术使用四个字母。这听上去似乎限制太严，但我们应该记住，计算机使用的二进制算术只用两个字母。单单一个人体细胞，它的信息储存量，就像三十卷的大英百科全书，可以装三四套而有余。

D　NA和RNA的演化变异可以用作一种分子钟。分子生物学家比较了现今活着的和已经灭绝的物种的遗传物体，发现D　NA和RNA在很长期间的突变率相当稳定。突变可以来自高能辐射，并且因为复制过程中有误抄。突变的结果是生出不同的如水蛭和地衣的物种。突变导致出一种演化钟，这个钟的“滴答”和突变率相应。这个钟可以用时期确定的化石来校准，校准以后，可以用来估计物种是在何时彼此分支的。它也被用来证明过，遗传密码不会老过38亿年左右。

在各种形式的生命里，基因语言指令着细胞机制，把蛋白原料的核酸连接起来，形成个别的蛋白。蛋白是另一群关键性的生物分子，上面我们已经遇到过，那里它们扮演的角色是生物学催化剂或者酵素。蛋白与核酸不同，蛋白没有自复制的能力，但是它们作为催化剂时高度的专一性，保证了自己和其祖先核酸之间有一个共生关系。这样，巨型的反馈环牵涉各种蛋白，因为蛋白对核酸的复制起催化作用，而核酸的复制对自身的产生又是少不了的。由于对化学反应动力学的贡献获得1967年诺贝尔奖金的哥廷根的马普生物物理化学研究所生化动力学系系主任艾根（Manfred Eigen）和维也纳的舒斯特（Peter Schuster），以及其他的合作者，尽力建造了一个理论框架，使此类的分子演化，借助于所谓的“超循环”——彼此关联的自催化反应的循环，能在一个糖和氨基酸的原始无机混合体中实现。他们从这个想法已经得到了一些预言，很可能在不久的未来就能得到实验的检验。

另一个关于生命如何开始的模型，是宾州大学和新墨西哥圣费研究所的考夫曼（Stuart Kauffman）发展的。该模型现由法尔摩（Doyne Farmer）、巴格利（Richard Bagley）、帕卡德（Norman Packard）等人继续探讨。这里设想的是一组遗传高分子或者蛋白高分子，它们能催化某些化学反应，使别的某些分子分裂、结合。简单的化学“养料”输入进去，然后转化为较复杂的分子。考夫曼与其同事指出，这样的系统是可以变为自我复制的。

以上的讨论让我们开始认识到，在什么样的情况之下，自复制式的化学反应可以得到发展。而让我们回想一下，“自复制”这性质是生命的主要特征之一。只要核酸和蛋白组成的分子集体受着非平衡的约束，各式各样的耗散式结构原则上便会出现；空间的结构，时间上的结构，时空中的结构，乃至混沌行为，都会出现；五花八门，也许就可以描述我们四周生物美不胜收的花样和形式。这个主题，将在下章详细讨论。

## 时间与创造

在本章开始时，我们曾让读者回忆一下，有些科学家的观点是认为时间之箭是幻觉。这些科学家，就像康德以及他以前的哲学家一样，认为时间之箭在热力学第二定律的出现，一如我们对时间流逝的印象，是和某些主观现象有关，或者跟大脑过程有关，而不是属于自然界的。

然而妙的是，我们把第二定律更仔细地考察以后就会发现，把时间之箭说成是主观性而置之一旁，反而会引起更严重的困难。看上去，要生命出现，不可逆的过程是少不了的。这个观点，我们将在下章探讨，我们将更仔细地考察非线性动力学在生物学中的应用。一种名叫黏菌的单细胞生物也好，心脏病发作时肌肉的行为也好，我们都找得到BZ反应的回声。在昆虫总数的涨落里，在关于性别的起源的看法之中，混沌也将要出现。

对“不可逆性是幻觉”学派的人来说，他们很难避免一个完全自相矛盾的情况。上面我们已经指出过，用含有时间之箭的方程，可以深刻地描述生命过程。如果该箭头是幻觉，我们就不得不说，各种生命的花样——包括我们自己在内，都是我们自己近似的结果。也许是时间之箭是如此深入我们的经验，以至我们忽视了它所占的中心地位。不过，一个科学理论如果容纳不下时间的这一面，要它来大规模描述真实世界时，肯定不会开花结果的。

# 第七章　时间之箭与生命之箭

没有活组织，就不可能有自觉和人类的创造。活而耗散的系统不会有，除非熵沿着它不可逆的通道在时间上前去。

皮考克（Arthur Peacocke）

《上帝与新生物学》

当你念这句话时，电火花在你的大脑中劈啪爆发。热带茂密雨林里，昆虫总数高涨低降。一个短鼻鳄卵中某点，注定了鳄鱼终身穿戴的一套花纹。在每个这样的生命图案之中，时间不停地滴答前去。

就像化学界有化学钟一样，生物界也有生物钟。这些钟的节奏，虽然彼此悬殊，然而对生命都是少不了的。神经细胞，一眨眼之间激发几千次。另种细胞里，物质的浓度以数秒长的周期起伏。组成人体心脏的细胞群体，每分钟跳动70次左右。而组成整个植物或动物的细胞总集，具有内在的、长达数年的成长发育周期。所有这些节奏归根到底都是由分子过程、生物化学过程所控制。要了解这些过程，我们可以运用与上章我们讨论贝鲁索夫-扎孛廷斯基（BZ）反应里时钟式振荡时所用的办法。唯一重要的不同就是我们现在考虑的体系是活体系。

在BZ反应中，自组织表现为化学活动的旋转螺旋，其中成千上万的分子齐步组成时间和空间的宏观结构。在生物界中，相应的组织过程则是个别细胞聚成多细胞有机体。自组织也包含如昆虫总数的涨落、人体心脏的跳动之类的有序现象。的确，整个的人体都可以看为一个在时间空间自我组织的复杂单元。

因此不足为怪，生物系统具有和BZ反应类似的内部反馈过程。酵素，在被身体制造出来以后，接着又参与随后的、和它本身制造有关的过程。该酵素可能激励，也可能压制细胞的机制。这种非线性过程的结果很难预测，因为当酵素的量在变的同时，制造酵素的规律也在变。但是生命本身就是一个高度非线性过程。载着反馈过程的蓝本的基因，同时也负责调制我们身体应该如何读、如何理解这同一基因。

生物界富有反馈过程，并且和化学界的BZ反应和布鲁塞尔振子一样，这些反馈过程能产生三种不同的自组织形式。时间上的组织是振荡；空间的组织是图案；两者的结合则是活动波浪在空间传播。关于生命是如何持续这问题，这三种自组织在一起，给我们提供了不少的见解。

这里必须认清，反馈有本质上不同的两种：正反馈和负反馈。正反馈增加系统中的产量，一种配料自催化自身的产生就是一个例子。负反馈则降低产量，例如控制暖气设备的恒温器。屋子里温度一旦低过某个预先定好的温度，暖气设备就打开，将室温提高到所要的水平。该水平一旦达到，设备就关上。然而不可避免地，房屋将会在设备关了以后继续加温，结果是使温度稍微走过头。同样地，当室温下降以后，当设备再开动时，室温会比预定的“最低点”少许低一点儿。这样，负反馈结果是建立一个循环，其中温度慢慢地上升，下降，再上升，再下降……。这本质上是使系统稳定化，此类反馈一般认为在人体中极为重要，例如血压的控制。纯粹正反馈一般认为对人体不这么重要，因为它会导致不稳定的行为。可是，对于动物总数在时间上的变化，正反馈是很重要的。至于正、负反馈同时存在，这种情况极为平常，身体中白血细胞的制造就是一个例子。

## 生物学混沌

前面考虑化学钟时，我们发现有化学混沌。现在我们考虑生命

钟，情况好像是一样。虽然关于生物混沌的研究仍属初步，它看上去对某些重要而往往是不良的现象，应该负有责任。混沌似乎发生在体内组织解体，身体处于非正常动力状态的时候。有人断言，心脏病暴发时，混沌便崭露头角，一如在某些疾病时好时坏的变化之中。另外有人推测，对混沌的了解将有利于用脑电活动分析来预报癫痫发作。它甚至是演化少不了的因素，不过，这些论调不宜过分相信。近年来世上无疑地对于混沌的存在有“赶时髦”现象：一方面报刊科学通讯员将它大肆宣传，另一方面科研者抓住它，用它来赢得名声与经济资助。鼓吹混沌的人认为：混沌到处都有。然而，对于许多生物学中的例子，可以说陪审团仍在考虑之中。就如第六章所说，我们应该辨别内禀于系统的决定性混沌和来于外界噪音的随机变化。区别两者，理论上有很好的检验方法，但是实际应用起来，并非直截了当。这是因为，如果没有详细数据，我们依靠不了现有检定混沌特征的方法，例如计算自称为奇异吸引子的分维数。这些吸引子，就是其中资料最齐备的例子，也还没有人证明过是满足严格数学定义的。

尽管如此，在生物过程每个层次上，从细胞内的事件到细胞与细胞之间的事件，从有机体体内事件到诸有机体之间的事件，似乎多少的确存在有混沌和组织并列。这些事件都是不可逆过程，而不可逆性同时含有混沌和组织的配方，犹如印度舞神湿婆，一手拿着破坏的火，一手敲着创造的鼓。本章将继续调查研究生命的这两方面，希望能彻底领会不可逆耗散式系统在自然界的重要性。

由于熵在增大，所有活体系都在耗散。这种体系也都是动态的，因为其中的过程极端地倾向于演化。用术语来说，我们现在是在研究远离平衡、非线性耗散系统中，时间上不可逆的演化。一个等价的名词术语就是“动力系统理论”。有些生物学家和数学模拟专家对“耗散”、“远离平衡”，以至“非线性”这类字眼感到恼火，把它们看成是别有用心的外行强加于现有学科之上、与题无关的商标。这种情绪虽然可以理解，总属可惜。对生物学的精明见解有些是从跨学科角度取得

的。基于不可逆过程的数学性质，我们将要考虑的每个实例都强调时间之箭的重要性，当然，一个理论仅仅应用成功，还不能证明它就是正确的。但我们相信，不可逆性在生物学中的重要性，已积累的证据是如此确凿，不容忽视。下面的实例，目的就是提出这些证据。

## 糖钟

我们身体中的细胞必须具有集中能源的手段。这样，细胞才可以与时间之箭交战，而不被时间之箭拖进热动平衡和死亡——时间之箭最原始的面貌。一群错综复杂的化学反应将食物的能量转化为细致的生命机制。植物动物只是在持续远离平衡的状态之下，才能产生生命必需的生理秩序。它们的细胞需要能量帮助消化，帮助合成生物化学剂。这样才能产生浓度梯度，肌肉才能收缩，体温才能保持，等等。

我们的家庭生活靠煤气或电。身体中的直接燃料叫“腺苷三磷酸”（ATP），是一种富有能量的关键生物分子。如要生命继续下去，ATP必须合成。ATP载能用的是一个高能化学键，它像一个压紧的弹簧，其中牵涉到化学家所谓的磷酸群，由四个氧原子围绕一个磷原子组成。ATP一旦失掉磷酸群，就退化为“腺苷二磷酸”（ADP）。反过来，ADP也可以经由磷酸化这一过程，进化为ATP。

绿色植物利用日光把A DP（和糖类）合成为ATP，此过程叫光合作用。而动物则用呼吸制造ATP。动物吸取糖类和脂肪，然后在它们细胞中名为粒线体的特殊操作单元，用呼吸从大气得来足够的氧气，燃烧这些材料。燃烧在一串连锁反应中进行，这些反应彼此策应好像手表中的齿轮。产生的废料就和一般燃烧的结果一样——水和二氧化碳。此种新陈代谢途径叫呼吸链。

然而，并非所有细胞都从阳光或氧气中取能。有些用呼吸链的二等代用品——酵解，使葡萄糖发酵，将葡萄糖分子一截为二，这样制

造的ATP要少得多。巴斯德1861年对酵母进行实验时，指出这些贫氧过程的效率远比富氧过程的效率低。正如一位幽默家所说：“这些实验证明了，没有空气，生命不是不行，只是太贵——就像今日纽约一样。”原始单细胞有机体，诸如也出现在酸乳酪中和食物中的酵母，即使没有空气，也能依靠酵解继续生存。蚝、绿海龟等生物也这样，它们大部时间生活在水底。甚至在人体中，酵解也起作用，尤其是在血液输送有限的地方，例如正在进行剧烈活动的肌肉。

在光合作用、呼吸链、酵解这三种能源中，生物化学钟一直滴答前去。我们了解最清楚的是酵母中的酵解。酵母是个别独立生活的细胞，有一种酵母生活在葡萄皮上面，使葡萄汁变成酒的就是它。发酵时产生的不仅是酒精（乙醇），还有别的产物，总数达500种左右，它们之间的比重决定酒的质量。此过程人们熟知，古代文明人都早已知道。根据布鲁塞尔自由大学的巴布罗延兹（Agnessa Babloyantz），“有关酵解最早的记载是埃及某座古墓墙上画的一套解说图，该墓属于法老突特摩西斯三世（公元前1505～前1450年）的酒窖管理官”。

酿酒酵母如何把糖转化成酒精，这项迷惑、陶醉了历代科学家的研究，是现代生物化学的一个前例。因为此项研究搞得如此深入彻底，我们对生物化学节奏图案的了解，在此情况下达到最高峰。事实上，到了1940年左右，酵解的整个新陈代谢途径全搞清楚了。1957年，兑森斯（L.N.Duysens）和阿米斯（J.Amesz）首次指出，酵解过程中，能量并不总是平稳地产生，有时能量以一定的节奏振荡，过程中各种中介物的浓度也同样地振荡，其中最重要不过的就是我们的富能老友ATP。

ATP浓度是否在时间上起伏，完全决定于当时周围糖和ADP的数量。振荡在生理调制中扮演何种角色，关键在此。细胞中如果只有少许A　TP（于是较多的ADP），酵解之门就打开，制造所需的ATP分子，细胞或许从其储备提取淀粉或糖原；如果细胞中ATP很丰富，例

如呼吸链工作一直很顺利，酵解通道就切断了。此调制过程叫做巴斯德效应，它实际上被单一种酶索所控制，这是一种大而复杂的生物蛋白分子，能加快某些特定的化学反应。

此酶素叫磷酸果糖激酶（简称PFK）。经过几百万年专门朝这项功能的演化，PFK每当ADP浓度足够高时就打开，每当ATP浓度够高时就关上。可是PFK是一种磷酸制剂，它利用AT P把磷酸群挂在糖分子上，从而使ATP转化为ADP。而ADP本身当然又是激活此酶素的因素，使它更快地工作。这反馈恰恰就是自组织所需要的那种自催化非线性行为。

目前有数种理论模型，目的是描述糖钟是怎么走时的。一般说来最成功的是布鲁塞尔自由大学的哥尔德贝特（Richard Goldbeter）和勒菲弗1972年提出、随后由哥尔德贝特及其他合作者所补充的模型。他们把问题中的所有枝节都删掉，只留下要点。开头是用12个连锁非线性微分方程来描述；最后只用两个这样的方程，把全力集中在酶素PFK和生物化学能源分子ATP。

如此极端简化的好处是：描述其中节奏的方程和描述布鲁塞尔振子的方程就很相似了。如上所述，它的化学钟性质可以表示为一个极限环。如果浓度适当，糖钟里ATP和ADP的含量便顺着一个圈，像跳华尔兹舞一样，变化周期大致一分钟，和实验值符合很好（图22）。这样，酵解节奏是生物界耗散结构，时间上自组织图案的第一个公认的实例。

有节奏的生物化学过程已发现许多，这些过程有的牵涉单独的酶素，例如自催化剂辣根过氧化物酶和乳酸过氧化物酶，有的牵涉到一群酶索。一如生物钟，这些振荡子对细胞内外信号的传送以及诸如胚胎细胞有的变为脑细胞、有的变为肝细胞的分化过程，都起作用。和前面讲过的情况一样，凡是秩序以可预测的循环形式出现的场合，混沌也就出现，其中振荡的频率幅度不停地做不可预测的变化。有规振

荡和混沌振荡，在植物用以转化日光为能的光合作用、细胞“发电厂”粒线体的呼吸之中，我们都不能忽视。

## 细胞分裂的万能钟

控制活体基本成长过程——细胞分裂的单独的钟，似乎也是自组织在主控。每当我们体内十兆细胞中有一个分裂时，就有新的遗传物质被制造、被分开、被隔离。细胞分裂调制得好，就得出诸如耳鼻之间的差异，调制不好就是肿瘤。体内细胞以各种各样的速率进行分裂。成熟的脑细胞根本不分裂。肝细胞每一两年分裂一次，而肠壁细胞一天就分裂两次。在细胞分裂过程中，我们再次同时遇到单向式时间和循环式时间。无论从哪一点来看，细胞分裂都是循环，因为刚分裂出来的细胞和其母体完全相似。细胞以一个固定的时间间隔分裂，每繁殖一次，细胞就打一下拍子。可是一系列这样的循环，产生出来的是线性时间，因为我们的细胞，多数都是命中注定了分裂繁殖若干次以后就要死亡。

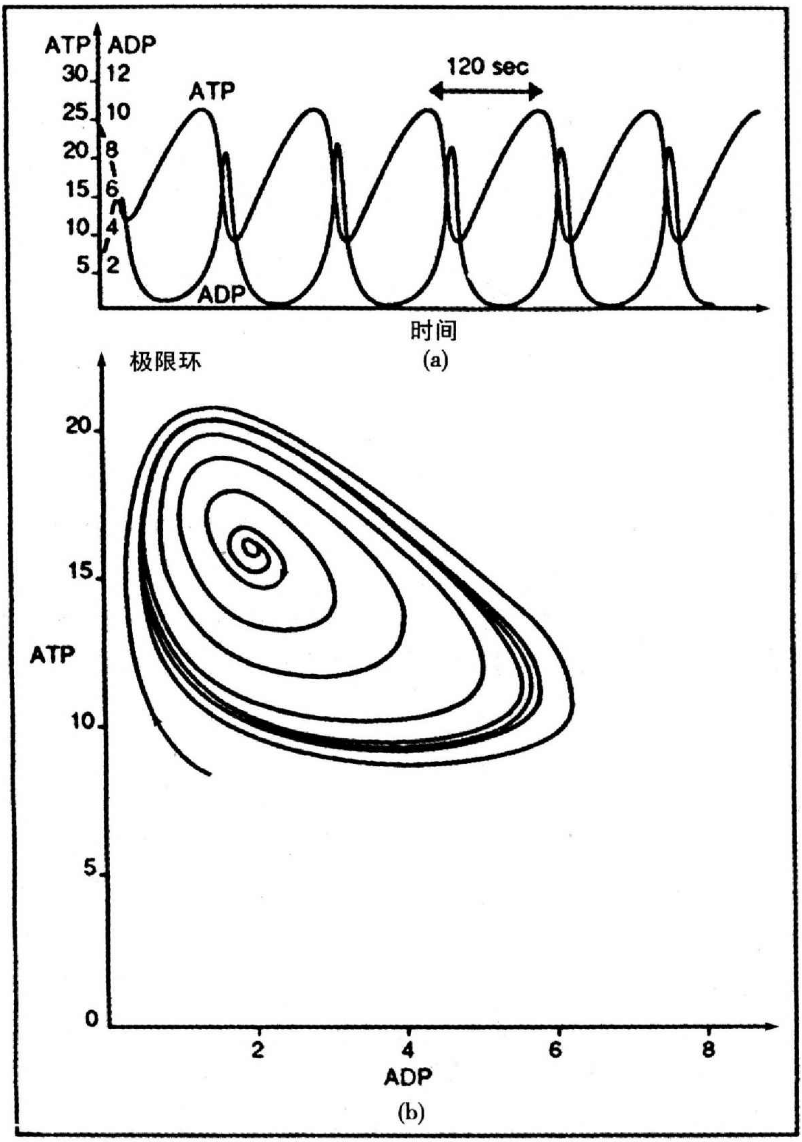



图22　（a）ADP浓度和ATP浓度的周期性振荡。（b）ADP~ATP相平面中的轨道逐渐演化成一个极限环。

谁都知道，我们体内细胞并非永远一直分裂繁殖下去。胎儿细胞分裂50次左右以后，就会死亡。一个40来岁人同样的细胞，分裂40

次以后就停止，而80来岁人的细胞，勉强可以分裂到30次。这次数和整个身体老化程度相应，寿命较短的动物，它们细胞分裂较少次数以后就停止了。未老先衰的沃纳症患者，他们的细胞死亡以前分裂次数特别少。

对于体内控制细胞分裂的基本时钟，近来取得了深刻了解。我们的身体从一个单细胞，经一系列分裂而成，每次一分为二。该过程严防意外，一步连一步，都经由许多因素预先安排好了。有的因素是细胞的内禀因素，例如细胞一般长到某种大小以后才进行分裂。有的因素则受周围环境的控制，好比说，要看细胞在整个活体中所处的部位。但是，究竟是何种分子使细胞分裂？对于左右这重要关节的遗传程序，我们能说些什么呢？

1980年后半期，研究者对这些问题，采取两种处理方法，从而取得了大步进展。一是科罗拉多州丹佛城的马劳（James Mallor）领导之下的一个研究小组和法国蒙沛黎诶分子生物学研究中心的皮卡（Andr Picard）、拉贝（Marcel Labb）、多雷（Marcel Dore）领导下的一个小组，认证了触发细胞分裂的分子。另一个是纳耳斯（Paul Nurse）领导的遗传学小组证明了这个重要的分子，对于从酶素到人基本都是一样。纳尔斯现在工作于牛津大学的皇家癌症研究基金会细胞循环小组。

细胞可以被看做微小的钟，摆动于两个状态之间（图23）。在一个状态下，细胞长大，分裂被抑制。在另一个状态下，虽然继续长大，细胞却进行分裂。细胞来回于这两个状态之间，一个来回就相当于钟的一次“滴答”。酵素的基本遗传程序中这部活钟的蓝本被纳尔斯与其同事发现了。他们第一步是把细胞分裂必需的一百多种基因认证出来。每种基因相应于一种对细胞分裂起作用的蛋白。可是光靠这点我们仍不知道分裂的机制。下一步就是要知道哪些蛋白重要。纳尔斯体会到，他如果找到一个基因能加强细胞的分裂，那个基因便相当于一个关键性的蛋白。他用钟的比喻来解释这道理。要钟走慢一点儿或

者要钟停不是难事，只要拿走某个钝齿，某个发条，或者其他一堆元件就行了。但只有使钟走快的操作才能揭露哪个元件是控制速度的关键部分。要找出分裂更快的变种很简单：一个细胞长得越快，它分裂时的体积就越小。纳尔斯和他的小组在爱丁堡大学找到了四个这样的酵母，起名叫“小变种”。其中之一最有意思，含有一个有缺陷的基因，名为cdc2（细胞分裂控制2），纳尔斯把它称作那年的“新闻变种”。它是一大族名为激酶的成员之一的遗传蓝本，这些激酶把别的蛋白打开、关上，办法是把一个带负电的磷酸群放在蛋白上面，使它们形状改变，从而改变它们的催化功能。看上去，细胞准备分裂的时候，此cdc2蛋白激酶打开许多分裂必需的蛋白。

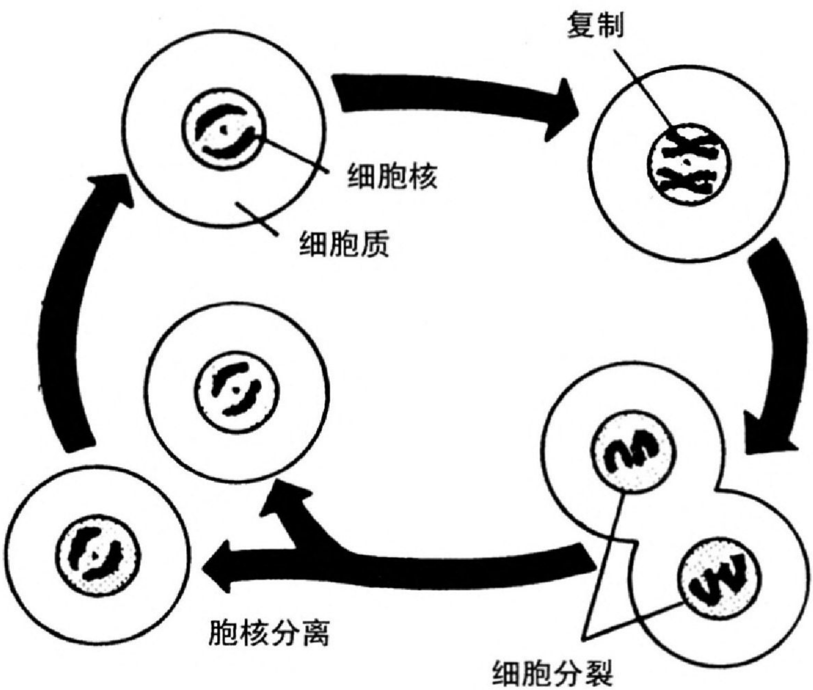



图23　细胞钟：细胞成长和分裂的各个阶段。

与这项工作相互呼应，当时还有人研究，是什么因素使未成熟的卵分裂。结果发现，如果把已经在分裂的细胞加在这样的卵里，卵就

会分裂。令人惊讶的是：所加的细胞可以来自别的生物，海星、人、酵母、“草履虫”都可以，青蛙、海星未成熟的卵都会被触发分裂。很明显地，控制细胞分裂的关键分子，从最原始的生物到演化最高的生物，都没有什么大不同。1988年，马劳在丹佛发现了这么一个分子。纳尔斯问他要不要把他净化过的蛋白和纳尔斯自己的酵母蛋白激酶对照一下。果然是同一个东西。这结果纳尔斯总括为“有了！成了！”（Eureka）。

牛津小组的李（Melanie Lee）做了一个漂亮的实验，证明控制人体细胞钟的也是同一种酵母蛋白。虽然细胞钟的详细机制仍属未知，但是很明显，管制细胞分裂的遗传程序，自从地球上有生命开始，亿万年来实质上没有多大变化：酵母和人有个共同的祖先，10亿年前活在地球上，用着和我们今日同样的办法控制细胞分裂，这是达尔文学说一个有力的证明。大自然的细胞分裂时钟，比作手表也好，座钟也好，控制机制都是一样，都是一个分子振荡子在调节细胞，让它来回在两个状态之间。这“同一种蛋白大家都用”的现象，在自然界屡见不鲜：每当大自然发现了一个有效方法去执行某个紧要任务，她总是尽可能重复使用这同一个机制。这种保守倾向不难理解：一个必不可少的酶索如果变异，形式如果稍有改变，那很可能就活不下去。它就会不再适合自组织所需的、错综复杂的生物化学反馈过程。

了解细胞循环如何调节，下一步是数学模拟。就如纳尔斯所说：“我们现在已经知道了谁在演戏，演的是什么。现在我们要知道的是：为什么某个角色在某个时刻上台。”首先，有一些问题必须解决。现有技术很难测量细胞中某个生物化学剂浓度在时间上的变化。分子生物学家在寻找关键分子时所用的手段，大半会破坏原有的空间时间组织。按照纳尔斯的说法，“我们想了解数据，但数据本身已经失掉活细胞动态描述必需的信息。”

细胞钟更多的详细情况得到以后，我们差不多就可以断言，这一个最基本的生命形式的秘密，迟早将会被动力系统学语言说穿。所有

必需的分子步骤都得连成一片，然后建立一个适当的非线性动力模型。美国理论生物学家泰森和考夫曼早年（1975）曾在这方面试过，并且自称得到了极限环的证据，当时极限环概念很时髦，就像今天的混沌一样。可是更详细的实验研究证明了，他们基于一个特殊模型的结论是错误的。按照亚利桑那大学温弗利的意见，这件事给我们的教训很清楚：这种非线性数学模型的运用，“必须逐次分别加以考虑，时髦概念的预设是用不着的。”

## 黏菌社交

让我们现在跨一步到生物界的次一层，不再讨论细胞里面的组织行为，而是讨论细胞与细胞之间的组织。像哺乳动物这样的高级生物，每个含有亿万个细胞，这些细胞在从卵到生殖的过程中，组织成一个极为精巧的结构。在此过程中的任何机制，我们对它的了解都不够给出一个说得过去的数学描述。尽管如此，用非线性动力学可以理解形态起源，这点似乎没有问题。这是依赖时间之箭的、最令人惊奇的创造过程之一。

一种叫黏菌的奇怪生物，它介于一堆单独的细胞和一个有机整体之间。它的学名叫Dictyostelium Discoideum（参见黑白插图），它妙在有时候是多细胞体（10万左右个细胞），而别的时候，如温弗利所说，“这些细胞独立地来来去去，像蚁群中的工蚁一般。它一如蚁群，是一个‘超有机体’，一个遗传一致的整体，由独立的、不自私的单元，为着集体利益组织在一起。”

黏菌细胞以细菌为食物。周围食物多的时候，每个单独细胞都尽情狂吃，以直接分裂方式繁殖。不用说，理想世界不能持久，这社会迟早要缺粮。到这时为止，诸细胞一直都不管彼此的存在，像孤独流浪人一样地行动。现在它们“注意”到彼此了。由于某些目前仍不清楚的理由，某些细胞开始活动，成为“领导”，有节奏地发出一种化学

剂，名为环腺苷单磷酸（cAMP）。此化学剂cANP在生物界极为普遍，作用是充当邻近细胞之间的信使。这个要大家团结组织的军号声，以每秒数微米（一微米等于一百万分之一米）的速度传播。

细胞一接到信号，便开始朝着领导细胞的方向蠕动，那也就是cAMP浓度高的方向。它们同时把信号放大、传递，形成反馈，提供非线性，使更多的细胞向领导集合。细胞以脉动波的形式向中心聚集。

只要看一看黏菌就会使我们觉得，这里的过程和化学钟里的大概差不多。书中图版部分中的照片显示，细胞群体聚集形成的同心圆波和螺旋波和出现在BZ反应中的螺旋波极为相像。这种波出现在心脏肌肉的节奏跳动之中（关于这点，本章下面还要详述），也出现在传染病菌中，乃至旋涡星系里的恒星形成波。

细胞一旦形成黏腻的一团，便开始分化，形成一个尖顶，从那里，cAMP不停溢出。整个一团组织成了一个光亮的、多细胞的“鼻涕虫”，有一个头，一个尾，向前蠕动，找光，找水。这群细胞形成这简单有机体，前后一共要数小时。它身长一二毫米，在脉动源尖头的领导下向前爬行。过后它竖立起来，形成一个杆子，杆子顶上托着含有孢子的囊袋；最后囊袋破裂，风把孢子吹往远处。如果孢子停在适当的地方，它们便会发芽，重新开始这奇怪生物的另一次循环。

这种行为里面令人注目的生物化学，使我们联想到上面描述糖钟里的酵解反应。为此蠕动物体提供信号的信使分子，cAMP，是从ATP经过一种名为腺苷酸环化酶的酶素的作用而产生的。就像酵解一样，这里也有反馈出现：细胞周围介质中已有的cAMP又激发腺苷酸环化酶，从ATP中产生更多的cAMP。这样便出现了自催化，而自催化是自组织不可缺少的因素。哥尔德贝特运用了他先前研究酵母细胞中酵解振荡时用的非线性分析，在极限环基础上详细地指出，cAMP如何每几分钟产生振荡。这是自组织行为的一个极好例子；再者，现

在也有了cAMP混沌振荡实例。在D.Discoideum的一个变种里，已观察到以cAMP振荡形式出现的时间上的混沌，以畸变枝干和结果体形式出现的空间的混乱，这些现象，如添加一种名为磷酸二酯酶的酶素就都可以恢复正常。

## 生物形态的起源

大自然具有多种机制，使细胞组织成为各种各样、令人眼花缭乱的形状。如用最简单的方法分类——按照体积的大小，人的位置便是在离最大顶端的十万分之一的地方。鸟、鼠、蜂或许看上去不大，其实它们也处于离顶端的1%之处。其余99%都是昆虫和小虫，它们的平均长度仅仅3毫米——“这样大小的东西，就是掉在汤里，我们也很少注意到”。要详细描述多样化配方，为期尚远。然而，人在相当大的程度上，受着和其他生物，诸如昆虫、蒲公英、黑猩猩同样遗传程序的控制，这一事实使我们继续有希望能对演变有个一般性的理解。

有些科学家只想对实际过程略加澄清。有的却想发现生物界的牛顿定律。20世纪早期，苏格兰圣安德鲁大学动物学教授达西·汤普逊（D'Arcy　Thompson）在他关于多细胞有机体成长的专著中写道，有生命物体的进化必须受几何所规定的框架的限制。今日的生物学家，许多都会认为这是条死胡同，认为该学科的大步进展不是出诸笔墨，而是来自实验室中的观察。

尽管如此，在生物学实验研究激增的同时，理论学者也在背后悄悄地致力于了解这些生命图案如何在时间上编织出来。从理论上我们不仅可以得到酵解和细胞钟所表现的时间上的图案，我们也可以同样取得诸如条纹、点子、山形纹等的空间图案，这些可以看做为驻波：虽然组成花案的彩色分子本身可能在做剧烈活动，整个图案是固定不动的。此种可能性早被杜灵在他1952年的一篇论文中认识到。关于形态起源——有机形式的发展，杜灵的看法近来有人研究是否可以用来

解释诸如斑马的条纹、金钱豹的点子以及身体各部分的初始分化。

要使一个图案从亿万个运动不已的分子混合体中出现，其所需要的配料在本书第六章讨论过。其一是扩散；化学浑汤不同的部分到底是经过扩散彼此才“交换信息”。另一个是化学反应。分子一面扩散一面起反应，反馈出现，于是花样产生。为了对这些空间结构做数学模拟，非线性动力学必须描述不可逆化学反应的速度，同时要顾及各种化学配料不同的扩散率。因此，杜灵的基本途径现今称为“反应-扩散理论”。第六章提到的“可激发性”可能也很重要。

生活在淡水里的水螅，有人说有100个，有人说有50个，有人说有9个头。虽然古希腊的大力神，借助于其驭者，终于结果了它的性命，它的一个低贱的远房亲戚现今仍然健在，并且在对形态发生的研究中做着贡献。从水螅身上截取一二毫米的一小段就可以长出整个的水螅。这惊人现象早在1744年就被出生于日内瓦的动物学家特兰布利（Abraham Trembley）注意到。它为形态起源的研究，提供了一个虽非典型、确是方便的系统。如果从水螅头附近取出一小片组织，放在身体的另一个部分，48小时以内就会长出一个新头；那个新头甚至可以拿走，它自己又会再长。“局部组织受到某个物体的感染了”，马普病毒研究所的马因哈特（Hans Meinhgrdt）这样写道。马因哈特对此形成图案的能力，在杜灵式反应-扩散理论的基础上，进行模拟。按他所说：“很有可能，在不久的将来，关于简单有机体成长的控制，水螅将会提供一幅颇为完整的图画。”

水螅中图案的形成似乎取决于两个因素：短程的化学激发（经由自催化）和远程的抑制。产生出的非线性导致许多有机体所共有的图案特点。典型情况是：一小片组织开始与其周围稍有不同，开始释放少量的“激发剂”，由于自催化，激发剂的浓度很快地增高。局部的高浓度触发抑制信号的制造，这是另一种生物分子，它散布到周围的组织，使别处不生产激发剂。这些所谓“形态子”的浓度分布轮廓，实际上告诉细胞它相对于特殊组织的位置，从而决定此细胞演化为头细胞

还是尾细胞。例如，要看是昆虫身体的哪一段，结果是腿还是触角。一般认为，激发和抑制不仅模拟初始图案的发展，并且对诸如鬃毛、毛发、羽毛、树叶重复式结构的间隔，也扮演重要角色。反应-扩散理论已被运用于四肢软骨的排列，羽毛鳞甲的分布，动物的体纹以及蝴蝶翅膀上的复杂花纹。是否成功，尚有争论。反应-扩散式方程可以产生各式各样不同的图案。杜灵的母亲写道："他给我看了一些（图案），问我它们像不像牛身上的斑点。它们的确很像，以至于现在我一看到牛就想起他的数学图案。"

动物的这种图案一般规定在壳或子宫里的胚胎上面。图案形成的精确时刻和当时胚胎的大小，是决定动物终身穿戴的花纹的关键因素。数学模型启示我们，为什么处于哺乳动物体积分布谱两端的象和老鼠，它们身上的颜色比较均匀一致，而不太大也不太小，像短鼻鳄鱼、金钱豹、斑马，它们身上的花纹就会很不寻常。此类模型表明，金钱豹的尾巴太细，装不了斑点，斑点都合并为条纹。的确，按照牛津大学数学生物学中心的詹姆士·默里（James　Murray）教授所说，数学模拟可以解释为什么"世上有身上是斑点、尾巴是条纹的动物，而没有这个样子反过来的动物"。

然而，用杜灵不稳定性来描述图案形成有个大问题，那就是：该不稳定性从来还没有在任何实验中真正出现过。对有些似乎控制发育的物质，有人建议它们是"形态子"：哈佛医学院的一对夫妇小组得到很好的证据，证明视黄醛酸会触发鸡胎中的一团细胞，使它们成长为腿或者翅膀；而坐落在曲宾根的马普生物演化学研究所的一个小组得到证据，证明果蝇胚胎的发育受bicoid蛋白梯度的影响。可是这类例子不符合杜灵关于花样产生的图像。其理由本书第六章已经略为提及。杜灵的反应-扩散理论无法描述自组织所有可能的机制。加州伯克利的吉姆·默里（Jim Murray）、奥斯特（George Oster）和其他人提出了一个与此有关但性质不同的处理方法。他们用一个类似的模型，但里面采用直接测量的生化量和细胞密度，他们叫这"力学一化学"方

法，因为细胞受到的化学和力学影响都考虑到了。此模型可以和已知量或实验可测量挂钩，它是否正确因而比较容易检验。默里使用了一个力学-化学模型，对短鼻鳄鱼身上花纹是在怀孕期哪个时刻规定在胚胎上，做了精确的预言，这个具体结果可以直接检验。此种措施进一步的运用将帮助我们了解，创伤如何自我治愈，正常发育期中四肢的软骨布局如何规划以及是何种因素导致先天性缺陷和畸形。

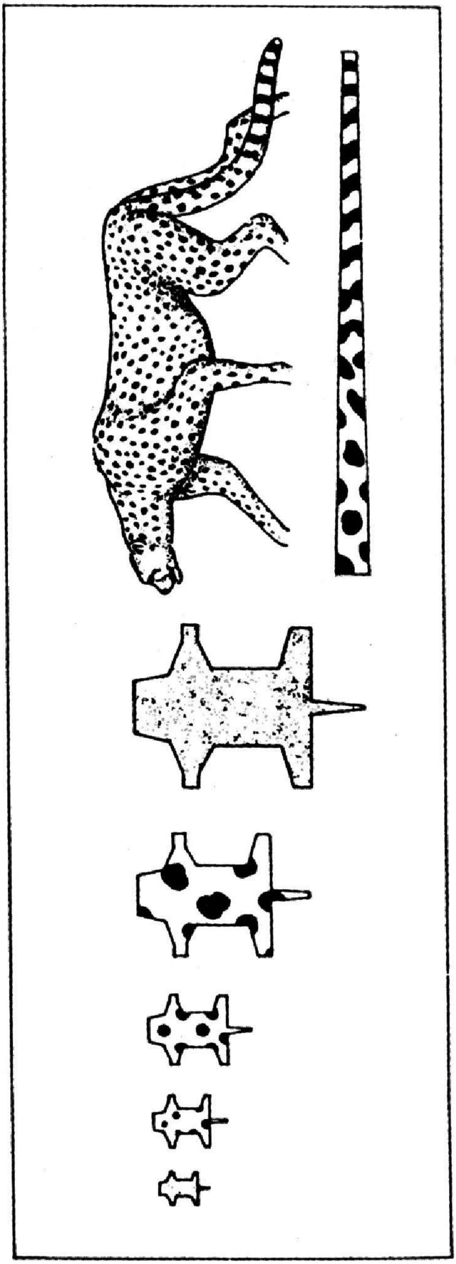


图24　在胚胎时期决定的动物身上的花纹和动物的大小有关。老鼠体积大小，因此多数身上没有斑点。体积大的动物，好比象。也倾向于全身一个颜色。用数学我们可以预言，面积不够大时，斑点就要变为条纹。猎豹下方的楔形图说明了猎豹身上和尾巴上花纹的不同。

## 跳动的心脏

形态起源是不可逆演化的一个极好例子，但它对我们心理起的作用远不及那台最了不起的生物钟——心脏。人的心脏一分钟跳动70次左右，一年4000万次，一生总共30亿次上下。这台充满活力的发动机，它所有的各种规则的节拍，都来自于不可逆过程。心脏不仅是生命的象征，它同时也是生命的一个共同极限：所有哺乳动物小如老鼠，大如海鲸，一生中心跳总数似乎都差不多，约20亿次。

心脏演化的前身是比它简单得多的生物泵。原始动物用这种生物泵，以蠕动的形式，使液体周流其身——今日人体中的大、小肠也是以蠕动形式运输食物。两个这样的管子经过演化彼此盘绕，长出极大的肌肉，最后形成我们今天四个心室的心脏。自从人类在地球上定居以来，有关心脏跳动的作品，从原始部落的歌唱，到历代的诗词，直到今日的流行歌曲，屡出不穷。现在轮到数学家和科学家显身手了。

在20世纪20年代，荷兰的两位学者范德泊（B.van der Pol）和范德马克（I.van der Mark）在这方面种下了种子，直到现在仍在给我们宝贵的启发。他们指出，只要调节模型中某些参数，便可以得到心脏规则节奏的各种不同的破坏——各式各样的心律不齐；每种心律不齐的出现都是经由非线性动力学中很熟悉的一个特点——分叉点。

何杰金（Sir Alan Lloyd Hodgkin）和赫克斯利（Sir Jul-ian Huxley）1952年的先驱工作，为生理系统的数学模拟带来了一大进步。他们研究的是鱿鱼的巨轴突——神经细胞传递冲动的线状伸延。这工作使他们获得诺贝尔奖。他们处理所得到的微分方程的方法，能给出一个定量的描述，现在经常被用来研究模拟心脏组织中电活动方程的性质。赫克斯利随后在1959年又指出，内禀于模型的是重复性活动，如果细胞缺钙，便会出现规则性的神经激发。

关于心脏动作的情况以及用于描述的非线性方程的数学性质，我

们现在知道的比以前多多了，对一大批心脏病的医疗，也取得了惊人的、不断的改善。例如，我们现在在心脏上可以置放一个器具，监视心脏的跳动，一出现可能致命的节奏，便摇它一下，使节奏恢复正常。很可能的是：不可逆非线性动力学这理论工具，对心律失调的诊断和治疗，迟早会有贡献。仅在美国，一年就有40万人由于心搏突然失调而死。这里节奏失调有时是跳动过慢，多半是心搏过速。虽然所谓过速会是快而有节奏的拍子，可是心电图会揭示，当心脏开始做纤维性颤动时，这拍子就退化为无规则的形式。这表示心脏细胞的一个非正常的时间空间组织。

许多生理节奏是由单个细胞或数个耦合在一起的细胞产生的。就心脏而言，我们现在知道心搏至少受到六种不同的指挥。一个特殊的肌肉组织，名叫普尔金耶纤维，它的细胞比心脏别的组织的细胞都大。虽然普尔金耶细胞并不直接提供心脏的自然节奏，它们把“窦结心搏引导”所发出的激发传运给心脏。这些引导细胞的节奏是由所谓“心搏波”触发，它们类似于BZ反应中的引导中心，那些中心可以用热铂丝稍碰溶液而建立，一粒灰尘或者盛溶液的碟子上一道划痕都可以触发它们。反过来，心脏病发作时也观测到与BZ反应中类似的螺旋波，围绕着一块健康的肌肉组织旋转。

牛津大学生理系的计算机里面跳动着一个活的心脏细胞——基于一个数学模型的数字式心脏细胞。这是诺布勒（Denis Noble）和他的小组设计的。诺布勒、弗郎西斯科（Dario di Francesco）和他们的合作者致力于建立一个数学模型，其中包括发生在细胞里的成千上万种化学作用。关于心脏跳动的原因，世界上的科学家一发现什么，他们就把它加进去。可是即使模拟一个细胞的动作也需要大量的计算机时间——要算一秒钟的心跳，诺布勒的计算机要运行100秒。因此要了解众多细胞的合作行为，会大大受到计算机时间的限制。诺布勒希望用明尼苏达大学的“连接机”（Connection Machine）来模拟高达50万个细胞的心脏组织。这是台威力巨大的计算机，具有1万个可以平行

计算的处理器。这类巨型数字运算可能帮助我们对不同的心搏分析加以比较，对一些问题不同的说法加以判断。例如，纤维性颤动有人说是起因于决定性混沌，有人说原因是旋转波的相空间奇点。

在本书写作的时候，细胞中的物理化学过程是用30多个联立耦合非线性微分方程来描述。这些过程中最重要的是让电信号闪进闪出细胞的渠道。这些信号由特殊蛋白传达，办法是把载电的离子搬来搬去。诺布勒的模型包括十多个这样的渠道以及另外一些在细胞内部和表面运输化学剂的过程。这里面最重要的是钙渠道，它触发ATP经过一串复杂的过程转化为心搏。与BZ反应一样，心脏也是一个耗散系统，因为物质进出细胞，要使细胞两边化学剂的梯度不断减低。

化学钟里面的耗散式结构是规则性的颜色变化或者美丽的彩色旋涡。心脏细胞里，众多蛋白纤维配合一致的动作可以看为其中的耗散式结构。为了维持这节拍，个别像钠泵这样的过程是不可缺少的。就像座钟钟锤渐渐地下落使钟不停，这种泵不停将钠离子输送过细胞膜，从而保证心脏远离平衡态。每次心跳，钙离子就涌入心房。跟钙渠道相互作用的不是别的，而是我们的老朋友cAMP。一如黏菌情况，cAMP和酶素腺苷酸环化酶并肩工作。在心脏里，这两位在控制钙渠道开关的反馈过程中，齐步动作。在单个细胞的层次上，涌入细胞的钙离子，运用一种分子“棘齿”机制，触发某种特殊蛋白的收缩。显而易见，钙进入心脏只是一半心搏。一次心搏完成以前，必须有另外一个运输机制，把钙取出来，放松细胞。这现象已经用特制的钙敏染料看到了。

得到的图像是依靠一群错综复杂的、自我组织的、不可逆反的相互作用的一个心脏——一群化学，蛋白、酶素信使的集体舞组成的每次心搏。诺布勒的模型，以其逼真的跳动，显示了心脏舞蹈可以用非线性微分方程加以数学描述：是一个极有考究的物理-化学“时钟反应”使我们心脏搏动。

## 动力性疾病

诺布勒的研究除告诉我们关于生物时间的一些情况外，还给心脏病的研究带来了好处。即使正常人，心跳率也有显著的起伏；当心脏钟里有多个彼此竞争的引导中心时，不正常的心搏常会出现，这情况有时就模糊地被说是“混沌性”。现在既然有了一个非线性动力心搏模型，我们就可以考查这种不规则性是来自决定性混沌，还是来自较传统的原因——随机噪音。要把两者从记录在心电图上的实验数据中分开，殊非易事。因此目前对于如何理解此类数据，意见不一。有些作者说混沌将会揭露心脏病暴发的隐秘，但这看法不一定恰当。例如，心室纤维性颤动是一种导致暴死的心搏失调，虽然临诊医师常说它不规则（意即“难以描述”），可是仍然看不出术语意义中的混沌[图25（a）]。

蒙特利尔麦吉耳大学的格拉斯（Leon Glass）、圭瓦拉（Michael Guevara）和希瑞尔（Alvin Shrier）对鸡胎心脏细胞进行的研究，可能是脏腑层次上生理混沌的资料最齐全的证据。心房细胞一些群体自发地、有规则地跳动着，给它们来个电震，下次的搏动便会提前或者推后。如果电震是周期性的，鸡心细胞便受两个频率的推动——细胞的内禀频率和外力的频率。这是能导致决定性混沌的经典情况。上章注意到，BZ反应中，改变化学配料增加的分量——相当于此处的电震，也同样导致外表上随机的行为。在酵解实验中，介质中糖的流动如有周期性变化，同样的现象也已观察到。类似地，取决于电震的频率，鸡心细胞在两次搏动之间的激发，也许是数目一定的，也许是混沌式的。

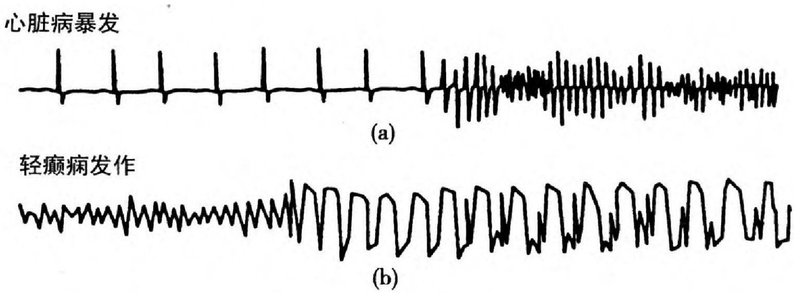



图25　两种情况之下的节奏变化。（a）心脏病暴发；（b）轻癫痫发作。[录自美国《科学》第243卷，第604页（1989）。]

批评者也许会说，鸡细胞的实验研究人为性很大；即使这里出现的是决定性混沌，它也不一定会自然地出现。波士顿的贝丝·以色列医院的哥尔德贝尔格（Ary　Goldberger）却认为会。哥尔德贝尔格在心电活动分析的基础上甚至宣称：健康人的心脏比病人的更为混沌。这表面奇怪的结论似乎违背常识，我们总以为许多身心功能的失调是由于健康机体节奏的瓦解。哥尔德贝尔格反而认为，混沌属于健康身体，而疾病跟此灵活性的丧失有关。对此观点，颇有争论。

温弗利采取了另一种立场。1987年，他猜想心脏肌肉如果在适当时刻受到适当大小的刺激，便会有出现旋转波式的心搏失调的危险。此后在心脏中诱发这种很快导致纤维性颤动的旋波的实验，支持了他这猜想。实验发现，如果电击施于心脏循环中某些特殊相位，顺、逆时针方向的旋转波都可以观测到。据温弗利说：“生物医学受物理理论引导的例子，少得可怜；这个实例真太令人满意了。”他对我们目前的理解，仍抱乐观态度。温弗利在他的书《时间破碎时》结论部分写道：“我首次遇到纤维性颤动这词是七年前，现在我还是不知道它的含义……它仍是谜。应该有人解这个谜了。”

心脏病暴发时规则性心搏的破坏，是格拉斯和麦齐（Mi-chael Mackey）所谓“动力性疾病”的一个例子。动力性疾病是由于身体正常节奏的移动而产生的。这定义巧妙地避免了混沌是好东西还是坏东西

的争论；只说病是因为节奏变了。这种动力性疾病，医生们都很熟悉，实例包括癫痫发作和各种呼吸失调，例如“澈恩-司托克斯呼吸”（Cheyne-Stokes breath-ing）——呼吸周期性地变快变慢，往往伴随心脏充血衰竭。不同的疾病像心脏病发作和癫痫，现在可以作为动力性疾病的例子进行比较。这从简称为EEG的脑电图[图25（b）]可以明显看出，脑电图显示的是活人脑中的电波，这些电波的来源是十万分之一伏特的电干扰。

脑电图是由放置在头皮上的电极记录得到；这是每个医院的常规工作。电极和状如蕈伞的大脑外皮调准。和心搏不同，一个正常健康人的脑电图是不规则的，并且是比较宁静的。当癫痫发作时，此活动起剧烈变化，此时病人会昏迷过去不省人事。中古时期，癫痫发作被认为是恶鬼附身，或者就是魔鬼本人。可是并不表现为脑神经失常。因为这时的脑电图中的电活动，虽然幅度比较大，节奏反而更有规则些。

巴布罗炎兹（Agnessa Babloyantz）与其布鲁塞尔自由大学的合作者，对癫痫发作时记录的脑电图，做了详细的分析。脑电图究竟能提供多少信息，不太清楚，因为用以记录的电极测量的是大脑颇大区域的平均电活动。尽管如此，这群科学家运用了耗散非线性动力学分析的标准技术，宣称在正常人和癫痫患者的脑中都找到了奇异吸引子——混沌的标志。这数学抽象的分维数大致表达混沌的程度：分维数越大，随机性也就越大。从脑电图的数据，他们计算得到的分维数是有变化的。对于一个活跃的正常头脑，他们未能把分维数算出来，他们只知道这时的分维数比熟睡时要高得多，而熟睡时的分维数少许高过4。但癫痫发作时，脑活动变为更有秩序，分维数几乎下降到2。

耗散式动力系统理论对体内节奏所提供的新认识，很可能加强目前医术的效果。一个引人入胜的课题是慢性骨髓白血病患者的生存率，这生存率50年来没有改进。有人认为，这是因为医生们没有充分注意到白细胞数本身也是有涨落的。格拉斯和麦齐相信，进一步了解

控制这些节奏的系统，将会导致更有效的医疗。

对于病人经受刺激以后有时出现的各种意外，动力学见解或许也会有帮助。这里所谓刺激可以是施诸心脏的电震失控，也可以是来自机械式血液净化器。最平常的情况是阿司匹林的经常服用或某种针药的经常注射时，进入人体的某种药剂。在此情况下，在身体自然节奏和治疗外加节奏之间，很不容易建立起稳定的关系。这告诉我们，应该寻找新途径来改善现有药品的服用，对此，有些药材公司已经开始考虑。在不久的将来，对生理节奏的理解会大大改进，医生们将能利用这一成果。有人甚至说，找出正确的服药时间，可以抑制规律性癫痫的发作，把刺激时间调好，可以抑制发抖，等等。

科学家们现在正在从非线性动力学中取得诸如此类的认识。谁想到过，从抽象理论的研究中，出现了医学上重要的进步？手里控制科研经费的政府官员、官僚、政客，他们大多数肯定没有这样想过。

## 性循环或硬激发

甚至性别，也被非线性动力学所表示的时间之箭控制。第五章给的分叉图理想化地显示出，当系统被驱赶到远离平衡态以后，所出现的可能性。在化学钟的情况下，出现了多种状态，这实际上意味着从一个具有规则颜色变化的振荡状态转变为一个非振荡状态。过渡方式之一，名叫“硬激发”，也叫做“次临界霍普夫分叉”。这意思是：当某个参数（例如某反应剂的浓度）增大时，以前一直没有的振荡或节奏突然发生。相信性交高潮就是生理方面的一个例子。

一个健康男性射精时，他骨盆底部肌肉的电活动如果记录下来，便可看到活动是爆发式的——是一串速射的神经冲动。此种事前没有显著的周期而突然发生的爆发，是和“硬激发”一致的，这点，读者或许会感兴趣，然而不会惊讶。已经发现的还有其他这种节奏突发的例

子，包括妇女停经期阵发性发热感。这种现象的反面也同样会发生，原有的振荡突然地消失。有人认为这就是婴儿忽然停止呼吸猝死的机制。

## 群体动态学

我们现在把注意力从生物体内的事件转向生物之间的事件。描述这些事件的数学，也同样具有时间之箭。这里一个很好的例子就是群体生长率，早在1220年，比萨城的雷欧纳多[又名菲波纳其（Fibonacci）]对此做了数学模拟。他给出一个极为可怕的预言，说一对兔子，如果让它们繁殖114代以后，它们子孙的体积将大过所知道的宇宙。随后有人指出，“远在此以前，地球就被埋在以超光速膨胀的兔子球体里面了。”

由于存在有掠夺、疾病、竞争、合作，大自然不会那么多产，那么不灵巧。兔子群体和吃兔子的狐狸群体，在同一个生态系统中进行错综复杂的演化。地球上的生命，竞争演化，彼此纠缠的程度，不亚于细胞中各种分子组成的交响乐。有人把细菌、哺乳动物、植物、鱼看做是一部周转日光、养料，全球性活机器的个别齿轮。19世纪，昆虫学家记录了寄生虫与其宿主之间有节奏的变化，开始意识到这种相互依存。大自然中的均衡必须调节得很准，否则某种生物会把它捕食的生物吃尽，然后自身饿死而绝种。

20世纪初，人们开始用数学模型来描述群体在时间上的变化。这类模型不可避免地含有时间之箭。结果得到是耦合的非线性微分方程，显示着比方说狐狸总数和兔子总数之间的相互依赖和它们各自在时间上的演化。群体动态学中最重要的成分是竞争（其实这也是演化的最重要成分，演化下面接着就讲）。兔子数目不太小时，兔子是狐狸的牺牲品，狐狸总数增加，兔子总数减少。可是一旦兔子太少，狐狸总数就要降低。这样一来，兔子总数将重新增高，整个循环又重复

一次。这种对有限资源的竞争造成一个起调节作用的反馈机制，该机制从数学观点来看，又是来自某些非线性。这相当于化学钟反应里的自催化。美国的劳特卡（Alfred Lotka）提出第一个振荡式的掠夺-牺牲模型，该模型由伏尔泰拉（Vito Volterra）独立地加以充实，伏尔泰拉是第二次世界大战前意大利科学界有影响的人物。此模型现在大家叫劳特卡-伏尔泰拉模型。

然而，这种时钟式的行为，这种规则性的起伏，在动物界中并不常见。这方面最详细的长期记录大概要算加拿大山猫（猞猁狲）总数的记录，原因是200多年来人们一直为其皮革猎捕山猫。从这些记录可以看到山猫总数有大幅度的起伏（图26）。这复杂的变化，许多人猜想是反映着山猫吃的动物——雪鞋野兔的总数的涨落，而这涨落又来自野兔食物数量的变化。可是劳特卡-伏尔泰拉模型不能用来描述这批数据，如果硬拿来用，就会得出野兔吃山猫的结论。

借助于非线性动力学，有人提出了另一种解释，其中只牵涉到山猫和野兔，而不管野兔食物的供给、天气的变化、疾病或其他外界因素。一个非线性动力学系统里。不规则起伏可能来自混沌。事实上，生物群数理论学者把决定性混沌这概念放在舞台中心，就是因为群数无规律的涨落是很普遍的、迫切需要解答的现象。“混沌”这词就是一位理论学者，马里兰大学的约克（J im Yor ke）在1974年创造的（研究报告发表于1975年）。

这种混沌现象的数学性质，梅尔堡（Myrberg）早在1962

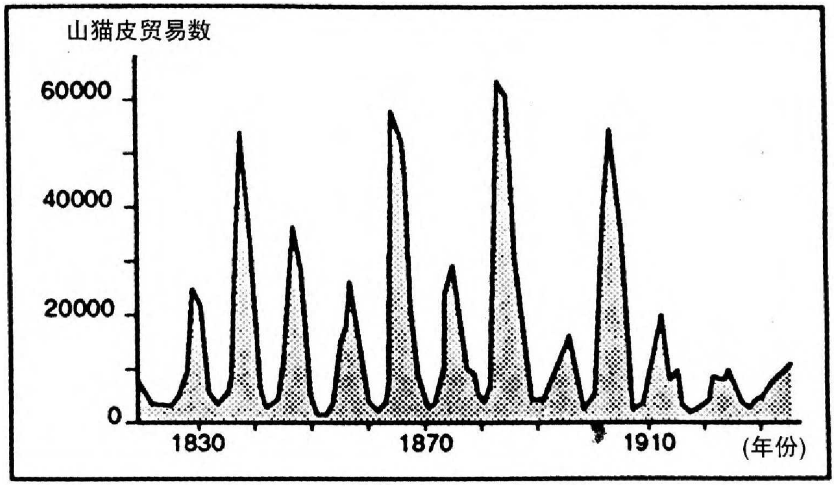



图26　加拿大山猫群数在1820～1930期间有显著的涨幅，每九年、十年达到高峰随着是迅速的下降。

年列出，随后又被其他人独立发现。梅（Robert May）无意中碰到非线性系统这些多得令人迷惑的可能性，1974年在美国的头号杂志“科学”上撰文，从而成为把它介绍给生物群体学者最早的作者之一。“现在回顾起来，很奇怪这种混沌动力学没有人更早注意到”，梅说。梅现在工作于牛津大学。在非线性动力学道路上，梅在试图大步行走之前，是缓慢地匍匐而行。他首先研究可能是最简单的群体动态模型——非线性“运筹方程”，此模型在母女不并存的情况下，对物种总数的变化，给出一个简要的描述。这里一个好例子是春天蜉出、秋天产卵后就死去的昆虫群体。这群体由于出生率的大小和对食物的竞争，会有多种不同的演化方式。出生率和死亡率可能相等，那时群体就处于一个定常状态，就不再变化。用动力系统理论的语言来说，这就属于一个定点吸引子。另一种情况是：群体总数可能出现有规则的跳动，在一些不同的定点吸引子之间跳来跳去，出现于两个、四个或其他数目的固定值之间。再有，总数可能似无规律地上下起伏，那时就相当于奇异吸引子所描述的混沌（图27）。

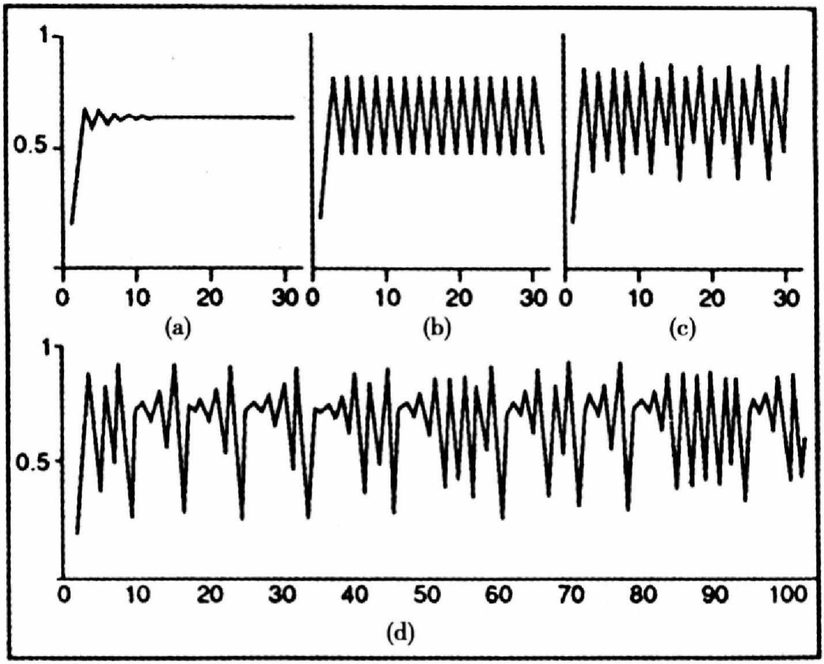



图27　同一模型（运筹图）中的各种行为；（a）稳恒状态；（b）2周期；（c）4周期；（d）混沌。[录自美国《科学》第243卷，311页（1989）。]

整个行为谱包含在单一方程之中。每种行为，只要换一下输入方程的一个参数，就可以算出来。变动此参数而得到的这假设的昆虫群体所有可能的选择，可以用我们熟悉的分叉图来说明。用计算机图解，可以把所有可能性画为一大幅奇怪的图案，图案的每一点代表一次计算的结果（参见彩色插图）。这些图是德国多特蒙德市马普营养生理学研究所的马尔库斯（Mario Markus）与海斯（Benno Hess）绘制的，每幅是数万次计算的结果，考验非线性方程的全部本领，结果就是这样。把方程中的参数稍变一点，得到的就是完全不同的另一景象。

梅的工作使人们惊奇：它说明，复杂性是一个非常简单的方程中的内禀属性。梅和他的同事奥斯特尔（Geroge　Oster）进一步研究运筹方程所属的方程族中别的成员的分叉图性质，——所谓“二次图”的

性质。其中有些与生物学似乎较有关连，在他们得到的分叉图的“树”中，他们在逐渐减小的尺度上发现有自相似性，从而揭示出一个奇异吸引子的分维几何。

这项工作的重要性——说明简单模型会有出人意料之外的结果，是梅在写一篇评论时忽然意识到的。梅讲道：“一旦有了二次图，道理连初中学生也能懂。只要在一个计算器反复运算。一个十二岁的孩子也会做。他能亲眼看到，一个简单、确定的规则会做出稀奇古怪的事。”

在刊登在英国《自然》杂志上的一篇极有影响的评论中，梅指出，群体动态学者所用的数学模型也出现在许多别的学科。除生物学中包括遗传学等不同的科目外，同样的方程也出现在经济学，那里它们被用来描述商业盛衰周期、商品供应与价格之间的关系，等等。在社会科学中，这些方程被用来描述谣言的传播。我们可以期望得到这些系统同样多种类的行为，从规律性振荡到时间上的混沌。梅的结论说：“不仅科研，即使在政治经济每日世界中，最好能有更多的人体会到，简单的非线性系统并不一定具有简单的动态性质。”如果主宰经济的是决定性混沌（而不是随机性混沌），财政部长们迟早是无能为力的。部长会说，调节某个参量，好比说利率，对经济会起长远作用。可是别忘记“蝴蝶效应”：按照动力系统理论，长远而论，一个领养老金的人把储蓄从银行取出来，也同样可能导致经济崩溃。

梅的论文使人们对生物群体的兴亡有了一个新看法。可是，我们如果把目光从抽象数学模型的冷静预言，转移到诸如上面讨论的山猫-野兔那种实际情况，我们就会发现，很难证明决定性混沌的确是观测到的毫无规则的群体兴亡的解释。的确，尽管15年来的努力，还没有任何人从生物群体研究中拿出一个得到公认的混沌实例来，其主要原因是数据不够充分，无法给出一个明确的答案。

为了克服这个困难，有人设想了一些吸引人的实验，来测验非线

性动力学的预言本领。其中一个是华盛顿大学应用数学家考特（Mark Kot）设计的，这个实验和“无机化学”中的贝鲁索夫-扎孛廷斯基反应十分相似。考特设计了一个捕掠一牺牲系统，其中捕掠者是一种名为原始动物的单细胞生物，牺牲品是某种细菌。亿万个这种有机体盛在一个开口的反应器中，细菌吃的食物不停流入，废物不停提走。在考特的数学模型中，食物输入如果稳定，这些微小有机体的群体动态就很简单。但食物的供应如果模仿季节而做周期性循环，预言混沌就会出现，就像第六章的BZ反应一样。目前进行的实验，目的就是寻找这种混沌。

另一些近来的理论研究是伦敦皇家学院的安德森（Roy Anderson）和梅合作的。他们相信在有艾滋病的免疫系统中，他们找到了混沌的证据。这种系统里可以说存在有一个捕掠-牺牲关系，艾滋病的来源——人体免疫缺乏病菌HIV“捕掠”白细胞和淋巴球T4，同时也刺激它们产生专门反抗HIV的免疫反应——B细胞与抗体。更广泛地，流行病学家现在用混沌来分析麻疹、腮腺炎之类的蔓延。阿利桑那大学的沙费尔（William Schaffer）是这方面混沌最热烈鼓吹者之一。沙费尔与其合作者分析了丹麦哥本哈根现有的流行病数据，自称对麻疹、腮腺炎、风疹都取得了混沌的证据，而鸡痘则有规则的周年循环。这些流行性疾病究竟是内在的混沌，还是外界随机影响的结果，现在仍不清楚。

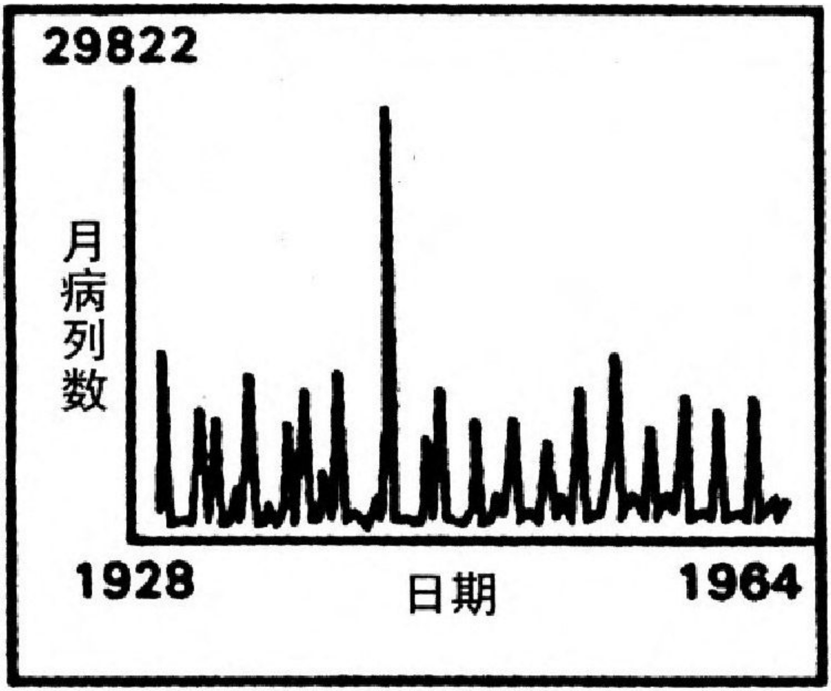



图28　有序和无序的混合体：1928～1964年纽约城中的麻疹病例。

## 进化与深时

现在让我们把话题转回到进化上。1859年11月24日达尔文《经由自然选择物种的起源》一书的出版，标志了一个智慧革命，其影响超出生物学，彻底改变了有头脑的人对世界的看法。达尔文的进化箭头，从简单的单细胞有机体指向复杂如人的生物，乍看上去是和波尔兹曼想解释的热力学箭头相反。但有一点比什么都重要，那就是：进化需要时间。地球存在的时间足够长了吗？

按照犹太-基督宗教的说法，地球创生后不到几天，人就出现了。17世纪，北爱尔兰阿尔玛市的大主教鄂修（James　Ussher）根据圣经，断言世界创生于纪元前4004年。可是当地质学家们认识到化石是一度生活于地球上的生物的石化遗体，而地质层是按时间顺序放置

的，他们隐约体会到“深时”。

地质深时的概念从17世纪起，地位渐渐提高，到了19世纪初期，地球年龄被估计在100万年到几十亿年之间。1863年，开尔文勋爵根据地球应该遵守的散热率，做了一个著名的估计。他的假设之一是：地球一度曾处于熔态，然后渐渐冷却；由此他得到结论，说地球支承生命不到十亿年。随后又有更精确的计算，把这可用的时间减到1000万年。

19世纪末放射性的发现，彻底解决了地质时间尺度问题。地球中地核地幔之间的温度梯度可能来自地壳中的放射性衰变，这样，开尔文的时间尺度就完全不对了。随后人们想到用放射性衰变的遗迹来测量地球本身的年龄。一个原子核自身的衰变率可以用来度量时间的流逝；这是一种很好的线型、非循环式时钟。衰变率一旦已知，放射性遗迹和其产品之比，可以用来估计地球年龄。这样，英国科学家斯特鲁特（John Wil-liam Strutt，即日后的瑞利男爵）在1905年给出20亿年的估计（用镭的衰变），而美国的波尔特伍德（Bertram Boltwood）估计是22亿年（用铀的衰变）。

深时的发现产生了极大影响，但这个概念很难掌握。用哈佛大学的古尔德的话来说：“对深时做个抽象的、理性的了解并不太难——我知道1亿是1后面放多少个零。但内心是否真正体会，那完全是另一回事。深时对我们说来太生疏了，我们只好打打比喻。我们高谈‘地质里’，人类历史只占其中最后的几寸地；或者阔论‘宇宙年’，人类只是在爆竹迎新年前几分钟才出场。”古尔德认为麦克菲（John McPhee）在他《盆地与山脉》一书中用的比喻最生动。麦克菲要我们把地球历史比做英国老长度单位——码，这是英国某个国王从鼻尖到伸臂后指尖的距离。国王手指甲锉一下就把整个人类历史锉掉了。深时——进化的奠基石，它的奥妙就在于此。

# 达尔文的理论

1809年，达尔文出生于英国西部希茹斯布利城。他在爱丁堡大学学医，当时他觉得地质课“枯燥到令人难以置信……唯一对我起的作用就是使我下决心，一生再也不读任何讲地质的书”。他随后考进剑桥的基督学院，准备将来做牧师，那时他“被裹入一群荡徒之中，其中包括一些下流青年”。的确，他父亲有一次对他说：“你什么也不管，整天就是打猎，赛狗，捉老鼠，你将给自己丢脸，给我们家门丢脸。”

1831年12月27日，达尔文乘官船“猎犬号”扬帆启程，于是开始了一个长达五年的环球航行，这航行将种下对有机世界一个新看法的种子。虽然达尔文此行是经由该船船长菲兹劳耶（Robert　Fitzroy）的邀请，他们两人之间关系并不好，再加上达尔文的晕船、生病，使达尔文麻烦重重。菲兹劳耶自信他会看相，能从一个人的面貌看出此人的性格：“像我这样的鼻子，他非常怀疑我会有参加此行的足够精力和毅力。”达尔文说。

1836年10月2日，达尔文回到英国海港法尔牟司。到此为止，他从一大批不同的地面、环境、动物、植物中，对他多年来猜疑的一个想法——动物植物经过长远时期能演化为别种动物植物，已积累了确凿的证据。但他直到20年后才发表这想法，而且还是因为一个英国年轻的自然学者华莱士（Alfred　Wallace）的一篇论文。华莱士在一个热带岛上害疟疾发高烧时，灵感一现，独立地想到自然选择的概念。一如达尔文，华莱士一直也在思考着马尔萨斯的人口论。华莱士写道：“我当时在呆想这理论对任何物种的作用，适者生存这意念忽然闪入我脑中。”达尔文大吃一惊，找了两个有影响力的朋友，组织讲座，同时宣读华莱士的论文和他本人1844年发表的一篇文章的摘录。达尔文的文章先读；虽然有人认为华莱士受欺，但华莱士本人却承认这位长辈比他早。

达尔文由于他的“猴子论”，成为漫画家挖苦的对象，且遭受科学家的严厉批评、宗教组织的强蛮攻击。他的论点是：自从地球上有生命以来，物种从较简单的形式演化成较复杂的形式，而自然选择——最适应环境的物种遗存下来，不适应的被淘汰，为此演化提供解释。这理论那时其实属于假设，而不能看做一个对世界的一成不变的描述：当时还没有任何确定性的证明，说这理论是对的。这理论的强处是它自称比其他任何理论更符合观测到的事实。以为达尔文式的演化一成不变是错误的；今日所有受尊重的进化派学者接受的都不是当初一字未动的原本。我们的知识一面在增加，这理论也就相应在演变，但就如梅所说：“所有这些研究都是在达尔文所树立的牢固框架中进行的。”

达尔文给地球上生命的多种多样形式下的配方，在现代原子、分子语言中，有了新的、更引人入胜的说法。在很大程度上，生命——至少在观念上，即使不谈它的细枝末节，可以从为其基础的DNA分子的性质中，得到理解。上章后部讨论生命起源时，已经遇到过这些分子。道金斯（Richard Dawkins）在他《自私的基因》一书中动人地描写了DNA自我复制的能力。我们都是“残存机器”，唯一的目的就是保护基因，就是要让同一条DNA链，也就是让这些巨分子所载的基因——决定我们面貌性格的蓝本，更多的自我复制。分子生物学的大步进展是这生命分子图案正确性的有力见证者。对生物内中的情况——小到每个分子，已取得了惊人的进步，得到的知识已用来设计了数代新药品，改变了生物的基因从而设计了新的生物，为诊断疾病和遗传性失调，设计了高度灵敏的检测手段。

可是，就如我们一再陈述，单纯的归纳主义也有缺点；如说今日的分子生物学有一点可批评，那就是：它复杂的枝节会把它的全面图像遮蔽了。对个别分子太关心，会使我们忽略了更重要的、分子之间的相互作用和时间上的合作行为。这种合作性的非线性效应，在远离平衡状态之下戏剧性地出现，代表着生命构成的一个不可缺少的因

素。

有人试图用热力学对演化做一宏观层次上的解释。布如克斯（D.Brooks）与怀利（E.Wiley）在他们颇有争议的《演化与熵》一书中，说了一句中肯的话："我们相信，大家都同意，有机体是处于远离平衡状态的耗散式结构。"对他们的方法说来不幸的是，他们混淆了不同的概念——第五章我们看到，自纽曼以来，信息论的概念和热力学的熵一再被人混为一谈。从热力学的确可以搬过来的是不可逆性的重要性，时间之箭的重要性。

进化过程是不可逆的，这看法最早提出者之一是法国出生的道罗（Louis Dollo 1857～1931），1893年他发表鼎鼎大名的"道罗定律"。可是按照进化论大师之一、梅纳德·史密斯（John Maynard Smith）的说法，自然选择并不意味着时间上的有向性。道金斯的著作里说道罗定律"常常跟进步是不可避免的这主观主义的无稽之谈混在一起，跟进化是违反热力学第二定律这无知的废话扯在一块"。并没有理由说进化过程中的一般倾向是不能倒转过来的，他说："如果在某段时期鹿角倾向于变大。很可能过一段时期以后，它又倾向于变小。道罗定律其实只是讲，同一条演化途径完全重走一次的统计概率是极小的……。道罗定律并没有什么神秘，它也不是我们要到自然界去'检验'的某种东西。它只是从概率的基本定理中得到的一个结果而已。"道罗定律中成问题的是它把不可逆性只放在"进行演化的系统"，而第二定律是讲整体的，既涉及此演化系统，也涉及此系统的周围环境（请回想第五章的讨论）。

在英国沃里克大学数学研究所曾任多年的所长，现为牛津哈特福特学院院长的塞曼（Christopher Zeeman），认为这两个概念永远不会扯在一起："从第二定律谈进化是不对的。模拟是在某个一定的，像物理这样的层次做的。下一个层次可能是化学，然后依次是巨分子结构、细胞、生物、生态学，最后是进化。每个层次开头都得用个不同的模拟。当你一口气又讲生命，又讲进化，又讲热力学第二定律

时，要知道它们之间根本没有什么关联。”按照塞曼的看法，生物学不像物理学，它是没有什么普遍理论的：“物理学中的普遍理论，对生物学说来无用，要不然就是引力会影响胚胎成长这类的话。你要引入一大部机械，但它给不出什么对生物学家有兴趣的预言。当生物学家们面对一大套数学时，他们就要问：‘这真正有用吗？’总的说来，他们不会浪费时间学这套数学。”奥斯特尔说得更尖锐：“以为热力学能告诉你生命如何演化，就等于以为把电视机放在热量计里就可以说明它如何工作一样。”

另一方面，许多作者，包括梅纳德·史密斯本人在内，都断言过越来越高的复杂性是生物演化的一个标志，这当然是指时间有向的一个过程。诺贝尔奖获得者卢利亚（Salvador Lu-ria）恰当地描述了进化的性质：“进化，一如历史，并不像掷钱币或者一场纸牌戏。它具有另一种不可缺少的特征——不可逆性。所有的未来都是现在的后裔，所有的现在都来自实际的过去，而不是来自可能的过去。人是现实的子孙，不是假设情况的子孙，而演化的现实——实际存在的生物，它们的范围只不过是过去所有可能的机会中很小的一个样本。”

美国一位古生物学家顾尔德（Stephen Gould）相信他在某些资料中找到了时间箭头的证据。他研究的是有同一个现在已经不复存在的祖先的一些生物，它们在时间上的分布。按照他的同事、弗吉尼亚科技学院的吉林斯基（Norman Gilinsky）所说：“历史看上去是不对称的。”顾尔德在演化树的枝干分布中找时间的箭头，其中树枝的粗细代表物种数目的多寡。他发现在物种数目增加的同时，有机蓝本的种类便在减少，枝干分布不对称，头轻脚重。对此他做如下解释：早先是试验阶段，品种数目大增，随后是规范阶段，数目减少，有机体衰亡。顾尔德认为这是化石资料中最显著的倾向。他写道：“这倾向比其他能从世系变化中而得到的结论，都更清楚地给时间确定了一个方向。这倾向很可能也反映着一个对自然系统如何演变的更一般、更基本的规则。”

最低最低限度，我们也可以说热力学的时间箭头和进化论的是相互不冲突的。既然化学、物理、数学越来越明显地赋予生物学活力，耗散式非线性动力学的概念也就很自然地渗入生物进化这课题。这是否对生物学带来好处，只有让时间来验证了。然而，对一些人来说，他们已经抵抗不了自组织和混沌的诱惑，他们在进化时间中已经看到了热力学第二定律中的同一个箭头。

生物化学家皮考克（Arthur　Peacocke）写道：“这样，宇宙表面上的腐朽、趋向无序的倾向，为新形式的产生提供了必需的、不可缺少的母体（妙词！）——经过老的死亡腐烂出现新的生命。”随后又写道：“普里高津和艾根与其合作者的研究揭示出，在导致有机结构出现的过程中，机遇和法则（即必然）之间，随机性与决定性之间的相互作用会是何等的微妙。他们的研究证明机遇和法则之间的相互作用其实具有时间上的创作性，因为是它们两者之间的结合让新的形式出现、演化——的确，自然选择看上去是投机取巧的。这机遇和法则之间的相互作用，其性质现在看起来是使有机结构不得不产生，不得不演化。”

主张生物学中运用热力学的人认为热力学提供的是整个图像的轮廓，而不是细枝末节，是可以借助动力系统理论把其他概念挂在上面的框架。情况和研究无机体中的自我组织时——BZ反应以及其他许多生物化学过程，完全类似：传统的热力学向不可逆非线性动力学让步。例如，皮考克与其他人一起，断言不可逆热力学“可以用来淘汰某些生物学模型，因为它们跟宏观物理定律相冲突，也可以容许一些模型，即使它对模型不能加以选择”。

## 演化的步伐

演化时间之箭是枝断箭：如果我们把所有现有的化石按照时间排列，它们并不像电影中连续的帧幅，每帧到下一帧的变化几乎看不出

来，而是似乎含有不连续的跳跃。上帝创生派和其他极端的原教旨派先用连续的演化箭头丑化了进化论，然后说这证明进化论是错误的。

化石资料当然是不完整的。但资料中这些空缺也可能对所谓“间断平衡”提供证据。“间断平衡”这词是艾尔德利基（Niles Eldridge）和顾尔德在1972年提出的。他们认为，根本不存在一个固定的演化率。物种倾向于在悠长的时间中保持稳定；它如要演化，演化就在很短暂的期间内进行。实际上，化石资料是由一层一层很厚的大陆土壤组成，每一层里面物种均匀分布，而层与层之间是物种突然变化的“面”。在这样一个“间断”点，一个物种一般演化成多个新种，如下面图29所示。

达尔文理论和艾尔德利基-顾尔德理论之间有矛盾吗？没有。塞曼证明了达尔文式的演化能解释间断平衡这现象。塞曼对这些演化跳跃的分析基于汤姆（Ren Thom）的“灾变理论”，该理论目的是：处理连续的“因”产生非连续的“果”的情况。达尔文式演化——随机性微小变异和自然选择，一翻译成最简单的数学形式，结果不是别的，就是图29所示的间断平衡。塞曼写道：“我把达尔文的话翻译成最简单的数学函数，然后环境的变化就预言这种间断平衡。”有趣的是：达尔文在其《物种的起源》一书中，唯一用以说明作者心意的插图，就是图29。

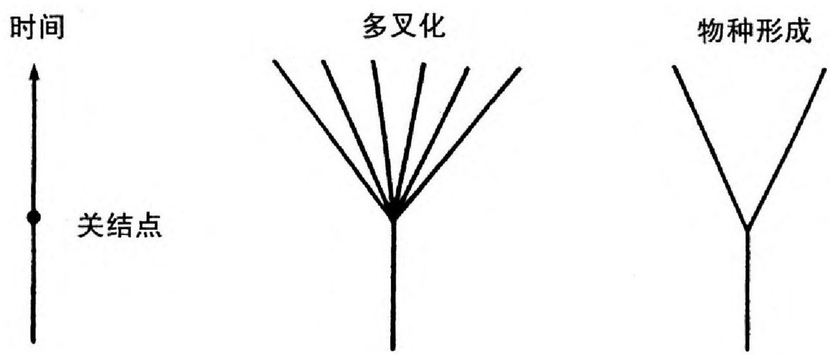



图29　多叉图、[录自E.塞曼《达尔文式演化的动态学》一书。]

尽管他的模型取得了成功，塞曼对数学在演化理论中可能扮演的角色，颇讲求实用。他相信，由于生物演化理论的很大一部分本质上始终将是描述性的，数学扮演定性的角色比扮演定量的角色更为可能。然而："正是在与直觉相违背、用语言难以说服，例如连续的前因产生非连续的后果这类的场合，数学可以期望起作用。"

## 混沌：性别存在的理由

演化论另一难题是性别，这问题可以用决定性混沌在时间之箭中的出现来对付。我们周围许多种生物是有性生殖。有些像细菌之类是无性生殖，那就是说，一个细菌不要另一个细菌帮忙，就能制造自己的复制品。还有别的，例如有翼小昆虫蓟马，它们两样都来。可是，为什么性别会演化出来？对进化最幼稚的看法是：基因是否被选，全看它的残存母体是否在为有限资源的战斗中占到上风。适应不够的就死亡。可是，梅纳德·史密斯在他关于进化与有性生殖的继续保存的一本专著中，宣称对性别的了解，目前仍缺少某个基本因素。使他感到困惑的是：短期看来，无性生殖更占优势。

有性生殖具有一些明显的缺点：找个对象会很麻烦；不同个体的基因聚集到同一个后裔，每个个别个体的基因总量只传了一半。这看上去似乎跟"自私基因"概念冲突。一个有性生殖的雌性动物平均每生一个雌性后裔时，无性生殖的动物已经生了两个；因此在两种都有的情况下，无性动物很快地会占上风。

可是性别也有它的优点，即使不算伍迪·阿仑（Woody Allen）所说的"这是我不笑而觉得最好玩的东西"。优点肯定是有的，不然我们也就不在这儿讨论这问题了。有性生殖比无性生殖更有效地把基因混合，使群体能有更多的遗传变异。这伸缩性的增大使群体更容易适应环境的变化。并且在短期内，有性生殖使群体更容易对付新兴的敌人，例如过滤性细菌，因为它使敌方更难预测。我们知道许多有性生

殖的群体，它们比无性群体更能抵抗寄生物的入侵；例如，引入智利的一种锈菌大大减少了一种无性生殖的黑莓（学名Rubus constrictus）的成长，而对有性生殖的黑莓（Rubus ulmifolius）则影响几无。

决定性混沌能帮我们了解，性别为什么在自然选择的过程中出现，使得我们更能对付寄生物。这至少是牛津大学动物系的哈密尔顿（Bill Hamilton）和他同事的观点。他们研究了性别在一个演化游戏中起的作用，其中相互竞争的宿主尽力设计对付寄生物的策略。在这计算机上进行的斗争中，哈密尔顿模拟了宿主和寄生物的遗传构造。简单说来，寄生物的目的是使自己与宿主的遗传构造相配合，从而扑向宿主，把它吃掉。游戏里，配合得越好，寄生物的“适应分数”就越高，它就越成功。配合完全，宿主遭殃。反过来，宿主尽量使自己的遗传构造和寄生物的“不配合”，使自己更容易避开寄生物。哈密尔顿的数学模型用13个耦合非线性方程来显示群体在时间上的演化。他认为，如对引入模型中的参数做现实的选择，结果将是严格意义下的决定性混沌。

按照哈密尔顿的看法，这里如果出现一个奇异吸引子，将对性别有利，虽然此点尚未确证。原因是：奇异吸引子带有混沌式的、不可预测的演化，有这一手，一个有性生物总比其寄生物占先一步。一个行为受其基因影响很大的生物（当然环境也起作用），混沌比光靠自然变异对演化提供更大的不可预测的余地。

哈密尔顿认为，混沌不是性别存在的必然原因；只是性别如果存在，经由自然选择的长期演化就更顺利。和单性生殖相比，两性生殖使生物更快地驶入混沌安全港。哈密尔顿并不是混沌的“传教士”，他谈起自己运用不可逆非线性动力学，态度颇为轻蔑；他说：“我的做法其实很笨。我只是把它放进计算机，然后看出些什么。”应该指出，像性别这本身惹人争论的课题，关于它为什么会演化出现，目前还有不少其他的建议。

# 生物界的节拍

在许多方面，地球上的生命像一支交响乐团，按照上天的节拍奏乐。太阳、月亮的运行，反映在昆虫总数的起伏之中，反映在全球有生命物体的活动之中。一切都是时间上的图案。

在用以描述这些图案的产生的理论框架里，时间总是向同一个方向走。在这枝时间之箭上刻着无止境的变化。有些图案人眼可以看到，例如金钱豹的点纹，戚纳巴鲁金兰的条纹。有些看不见，例如心脏的跳动，神经细胞的激发，细胞分裂的不断进行。这些时间上的图案不仅是生命的一部分，而且是生命的基础。即使表面看上去是偶然性的过程，诸如疾病的传染，鱼群总数的涨落等，里面可能都隐藏着秩序。不可逆非线性动力学里面包含既有自组织的配方，也有动力学混沌的配方。

宇宙在时间上普遍地、单向地前去，朝着一个最大熵的可能状态，在这过程中，滔滔涌出细巧有序而瞬息即逝的生命图案。否定这时间之箭，对诸如相对论、量子力学、经典力学等领域说来，或许是方便的——话说回来，这些理论适当运用在别的场合是相当成功的。但也许这些巍然的理论建筑未能说明事物的全部真相。

法国理论物理学家德·艾斯帕纳（Bernard d'Espagnat）曾说："从诸如灾变、耗散式结构等概念，到对有机物体的演化一个真正普遍的了解，这中间仍是相当大的一步。……但是这方面确有进步，这是不可否认的。希望将有更多的进展，早晚会渐渐看到出现一套理论，它非但不把'活力'归并于某些平常的机制，并且要把'活力'深奥的美，更清楚地揭示给我们看，就像经典天文学把宇宙的美揭示给我们先人一样。"

# 第八章　统一的时间景象

学说之间冲突，并不是灾害，是机会。

——怀特海（Alfred North Whitehead）

《科学与现代世界》

即使最顽固的怀疑者，经过上面这一趟科学之旅，也该相信不可逆动力学的力量了。我们用了基于时间之箭的分析，对一大批现象，从生命的出现到豹皮花纹的产生，提供了不得不接受的解释。如果我们说时间之箭是幻觉，把它放在一边，那我们得到的所有见解，都得放弃了。这牺牲可太大了；而我们换得了什么呢？一个充满各种荒诞无稽现象的世界观——碗里的汤自动热起来，台球神秘地从球袋里蹦出来。时间之箭的客观存在是一个不可否定的概念。如果这需要我们对某些习用的科学概念进行翻修，那也只好这样做。本章是全书最后的主要一章，这里把上面所遇到的、时间之箭是真实的各种暗示，汇集在一起。我们将把这些暗示跟近来对混沌的想法相结合，这样一做，我们将揭示时间的确具有一个能统一我们个人经验和科学经验的方向。

需要的是另一种描述，其中未来不是被现在或过去唯一地固定下来。严格的决定性论必须推翻；取而代之的是如下一个世界观：它和我们对世界的经验是一致的，它里面的未来是开放的，那个未来具有真正的演化和创新，能产生我们在大自然中看到的各种美丽图案，从黏菌的蠕动到错综复杂的全球天气系统，乃至宇宙本身由以产生的过程。此新观点真正地综合两个相反的、不可或缺的概念——机遇（概

率）和必然（决定性）。此观点赋予时间之箭十足的客观存在，使我们能理解实际世界中，不管是我们看得到的还是看不到的，已经发生、正在发生、将要发生的成千上万的过程。严格决定式图像摒弃以后，拉普拉斯式的机械宇宙与自由意志之间的冲突也同时丢掉。此看法要求完全接受热力学第二定律，承认它是时间有向性的一个表达，把它从头就放在我们的描述里面。现在先让我们回顾一下我们的出发点。

## 时间的问题

在我们寻找时间之箭的过程中，我们调查了现代科学中所有的主要学说。我们看到，牛顿力学、爱因斯坦力学、量子力学这些所谓“基本”理论，都否定了时间具有方向。我们看到，决定论和因果论与这些理论的可逆性紧密相连。爱因斯坦建造他十分成功的引力的几何描述，动机其实就是出于他对因果关系的第一位置持有根深蒂固的信仰。可是在这种决定性的理论中，时间被降到次要地位：不管时间朝哪个方向走，整个的未来，整个的过去，都包含在现在之中——这三者在某种意义下，只是同一整体的几个不同方面而已。

因为这些方程没有内在的时间箭头，没有理由取一个方向而不取和它相反的方向。但是情况比这还更差：时间不仅无向，它必须循环，“历史”必须按照庞加莱回归重复不已。就像圆是没有尽头的，这永恒的回返不允许时间有起点或终点。既然时间没有任何一点是和其他点不同的，平衡——热力学用以描述所有过程辗转而停止的时间演化终点，这观念也就站不住脚了。的确，如果这些理论正确，我们就得向热力学所有的概念告别，包括熵在内，而熵是科学中唯一出现的、对我们所感觉到的时间给出一个中肯描述的量。

如果我们追随一些理论物理学家，说世界根本是遵守决定性的、可逆性的定律，那我们就得说，我们自身的存在和我们所有的行动都

可以追溯到产生宇宙的初始条件或边界条件。在时间的黎明，这些条件就规定好了，未来行为应该遵从牛顿定律的这一指令，也已执行过了。或许是上帝选择了这些条件——他老人家亲自安排了大爆炸，使得生命和人类能够出现。在有些方面，剑桥前数学物理学家泊尔金洪（John Polk-inghorne）还要更进一步。按照他的看法，上帝不仅点燃导火线，然后站开；他老人家的任务是存在于宇宙万物之中，每个时刻，都是他老人家在保证自然规律的有效，在保证这些规律永不变质，在使台上的戏唱下去。即使如此，时间的实际方向不能用时间对称的力学来解释：它仍是上帝的、或者某个别的初始条件选择者的另外一个选择。初始条件预先就已假定了时间有方向。初始条件是额外的选择，与采用它们的理论的性质无关。本章结束以前我们将指出，牛顿决定论这水晶球破裂得很严重：这些条件，即使原则上也不能严格得知。因此，我们必须对力学中演化的意义重新加以考虑。

所有这些“无时间的”理论中，量子力学最为神秘：对它的理解和差不多所有我们对世界公认的看法都背道而驰。然而对理论物理学家来说，它是一个上好的工具，因为它对原子或更小尺度上的实验给出了预言其结果的数学方法。量子力学这我们现有的最好的物质理论，只要不想知道它的意义，完全可以轻松地赏玩。但想理解这理论是困难的，然而就是这些困难含有关于时间箭头的一个重要暗示。量子力学的中心难点在于测量，在于微观的原子分子世界和宏观世界碰头、测量器指针颤动、荧光屏闪光那个时刻。这研究世界必需的基本操作是不可逆的。上面第四章描述过，按照正统哥本哈根解释，一个诸如原子的微观系统，正是记录它行为这一操作，使其波函数坍缩，从而产生一个特定的结果（波函数是包含所有有关该系统信息的量子力学量）。这坍缩是不可逆的：它处于可逆薛定谔方程的框架之外。

前面曾经提过，彭罗斯正在追求宇宙学中的“圣杯”之一——在一个完全量子化的引力理论里面，把爱因斯坦的相对论和量子力学统一起来。理论物理学家多数认为彭罗斯的做法很不寻常，因为他承认了

把时间不对称性包括在宇宙基本特性之中的重要性。彭罗斯式的统一将有几个额外的好处：它将显含一个宇宙时间箭头，从而说明不可逆的波函数坍缩；它将清除广义相对论中令人为难的奇点；它将说明大爆炸的几率极小的初始条件（这样，不用初始条件来解释时间之箭，而反过来用时间之箭来解释这些初始条件）。这做法，一如其他宇宙诞生于无的尝试性想法，或许不需要什么神来开动万物。不过量子力学和相对论一个自洽的融合至今仍是未中之的。

在短期间内，物理学家最好还是放弃这些主意——别的方法可能更实惠些。实际上，物理学家都是实用主义者，绝大多数根本不管时间提出的问题。他们都乐于采取兰姆（Charles Lamb）的观点："时间和空间比什么都更使我烦恼；它们比什么都更不使我烦恼，因为我从来不想这些玩意儿。"天文学家巴罗在他《世界里面的世界》一书的结尾，写了他对这种态度的不满："对如何理解量子力学这问题，一个通常的反应是物理学家的反应，说量子力学行就可以了，别的就甭管了；说量子力学究竟是什么意思，我们搞物理的不必去伤脑筋。可是这种态度，用在别处，是不受欢迎的。如果一个学生问二次方程怎么解，说他只要知道给出解的公式，不想知道其所以然是怎么得到的，我们对这学生的印象一定不会太好。要知道科学研究就是基于'行就可以'这态度的否定。"

在少许对时间方向有兴趣的物理学家之中，我们看到许多人为了克服量子力学的测量困难，说波函数坍缩实际上并不发生。他们说做测量时，只是我们的见识在起变化。波函数的变形并不是指真实世界而是指我们的意念：不可逆性是出于我们对程序的干预。这样，"误解"、"主观性"便成为对此问题有限答案中常见的托词。许多科学研究者不承认这是发展新而有益的学说的机会，反而退到主观主义的说法，说不可逆性是幻觉。类似地，热力学第二定律被看做是麻烦，而不是自然界一个不可违背的事实。第五章讲到过，有人想用粗粒化把第二定律引入然后把它巧辩过去，可是这里往往有不对的地方。另一

个典型的循环论证的说法是：只是因为宇宙的初始状态具有变化的潜在能力，所以才存在有第二定律。

“时间之箭是幻觉”这论调引出一串高度毁灭性的推理。热力学——尤其是第二定律只是一个近似，只是我们自身的不足或“错误”的后果。再进一步，像生死这种明明是单向的过程，它们看上去的不可逆，只是因为我们无知，没有看到真正的时间对称的缘故。主观主义学派的知识论推到极点就是唯我论——自我存在是唯一的真实。这立场逻辑上是没有争论余地的，有少许科学家采取这个立场，大概就是为了避免不可逆性、测量等问题。考虑到科学为一个“外在”的客观存在所搜集的大量证据，看上去唯我论不太可能是正确的。再说，唯我主义者往往并不勇于坚持自己的信仰：他们一有了孩子，还是照样搞人寿保险。

事实上，就如上两章已经强调过的，时间箭头是增广知识的工具，不是用以掩饰无知的手法。我们周围看到的图案，从松果的朴素几何到猫身上的华丽花纹，原则上都可以从嵌有时间箭头的方程得到解释。对于远离热力平衡的场合，第二定律中祀奉的不可逆性原则将导致自组织过程，从这些过程我们可以理解自然界的各种有序结构。的确，如果没有热力学第二定律所坚持的不可逆性，预期生命就不会在地球上出现，也不会有表征活东西特性的各种时间上和空间的行为。只是因为有不可逆过程，我们才懂得我们自己的存在。

潜伏在不可逆性佯谬之中的大革命，感觉到它萌动的科学家为数极少，这或许不足为奇。学术界有越来越强的压力，要大家做专门研究，发表文章，从树林中找出个别的树，结果是科学文献指数式地增长，理解逐渐被牺牲在计算圣坛上。我们的前面其实是一片广阔的处女地，无数丰富多彩的可能性，仍待我们去探索。法国数学家、费尔德奖获得者汤姆在他《结构稳定性与形态发生学》一书中，如下呼吁：在这世界上如此多的学者忙于计算的时候，是不是也应该让能做梦的人做做梦？

时间新颖的可能性开始有人探讨了，特别是布鲁塞尔的普里高津学派。他们问：如何把时间各种各样的意义关联起来——动力学中运动的时间；热力学中不可逆性的时间；历史，生物学，社会学中的时间。普里高津写道：“这明显不是易事。然而我们生活在单一宇宙之中。为了对我们为其一部分的世界取得一个首尾一贯的看法，必须找出从一种描述转到另一种描述的办法。”我们无须对第二定律与可逆式动力学的对垒局面感到无能为力而袖手旁观。我们可以采取另种观点，它并不难以置信，它把第二定律不当为近似，而把它当做基础。普里高津与其小组提倡的这个途径，是基于对微观世界一个彻底的重新估价，它来自近来的一个认识，即动力学混沌除了最理想化的情况以外，其实到处都是。虽然动力学与第二定律永远不会彼此归化，它们看上去都像是自然界的内禀元素，它们之间的关系使人想到量子力学里的波-粒子二重性。

## 决定论放手了

在我们对时间箭头的搜寻中，我们先找一下牛顿，采用第二章所述的他对微观世界的描述。当然，这样做对寻找我们的对象不是有利的。牛顿定律说，信息如果充足，就可以决定任何系统的过去和未来。好比说，在一个罐子里不停碰撞的成亿成万的分子，如果它们在某个时刻的位置和速度完全知道，就可以预言、倒算它们在整个时间上的行为。庞加莱进一步又证明了这些分子将要在漫长的时期中不断重复它们的运动。在罐中分子乱动的运动之中，时间之箭丢失了。根据拍摄这分子运动的影片，即使运动可以看得见，谁也无法说气体是否达到了平衡。可是世界并非如此简单：牛顿力学的缺陷可以用来帮助我们揭示时间在微观层次中的方向。

在早期，当牛顿一给出他的运动方程，人们都以为预测苹果或行星的运动，不过只是应用数学中的练习。要知道月球如何绕地运行，只要把适当的数据放进牛顿的微分方程——月球在某个时刻的位置和

速度，经过一番计算结果就自动掉出来。此后200年，聪明的数学家和理论物理学家就一直把精力花在把这方程应用到越来越复杂的场合，然后寻找这方程的严格解。

碰来碰去的台球，沿着无摩擦力钟摆滑下的老鼠，要用牛顿方程来发现它们的过去未来，就得找出所谓的运动积分。这数学程序成功与否，当时人们以为全看是否能为台球或老鼠写出数学表达式。人们以为所有机械式描述都是可积分的，也就是说，都是能（从运动积分）得到严格解的。从来没有人停下来想一想，或许不可能严格地解这些方程。直到庞加莱出场。

庞加莱眼睛近视，心不在焉，举止笨拙。他同时代的人和他开玩笑，说他是两只手同样灵巧的人，因为实际上他右手跟他左手同样不灵活。的确，他在学校的成绩以绘画得零分而有名。可是，他的数学本领远远补偿了这些短处。1872年他学校的老师利尔德（Elliotl Liard）写道：“我南塞班上有个数学怪物，叫亨利·庞加莱。”

次年，这怪物发表了他第一篇论文，登在《数学新年报》上，当时他19岁。到他1913年去世的时候，他一共写了30多本书，500篇专业论文。他大概是数学史上最后的多面手大师，在学城索朋教学时，从一个科目跳到另个科目。对光学、电、弹性、热力学、量子论、相对论、宇宙学，庞加莱都做了贡献。他的一个同事说他是个“征服者”，不是个“殖民者”，因为随便他研究什么，他总引入新的思想。庞加莱也是科学普及大师。在他《科学的价值》一书中，他说：“如果大自然不美，那它就不值得认识，而大自然如果不值得认识，生命也就不值得了。”他被荣选为法国研究所文学部成员。

1889年，庞加莱震惊了科学界。因为他证明了，连只有三个成员的系统，例如太阳、地球和月球组成的系统，分析它们的运动，都会发现是个根本不可积分的系统。这是用术语的说法，其实就是指数学分析无法给出一个精确解。多过三个成员，更不用说一个气体中成亿

成兆的分子，要描述它们的运动，更是困难。这是牛顿水晶球的第一个裂缝。这限制有力地提醒我们，简并派想把一切尽可能地化简是危险的。如果把注意力全部集中在过度简化、能被数学征服的模型，便会有忽略真实世界整个丰富内涵的危险；特别是剥去一层层以为是模糊的现象而揭露内中的“基本”性质，这做法会使我们失掉时间真正的精华。

追求单纯是科学家一再堕入的陷井。牛顿写道：“大自然喜爱单纯，不爱过多因素的繁华。”可是对想象中最简单的情况，庞加莱把牛顿的数学已弄得无能为力。庞加莱写道：“一个世纪以前，有人在光天化日之下宣布了大自然喜爱单纯，可是此后不止一次，我们发现大自然并非如此。”在另一学科中，热力学的伟大先驱吉布斯坚持道：“对任何一门学问，理论研究的目的之一在于寻找一个观点，从那里整个学科看上去最为简单。”如果把吉布斯发展的平衡态热力学和他所忽略的、本质上更复杂的、跟实际关系大得多的非平衡热力学对比一下，我们就可以看出，吉布斯要求的“简单”是有缺点的。在所有情况之下，我们应该牢记剑桥哲学家兼数学家惠特海德的训诫，“追求单纯，怀疑单纯。”

## 遍历性

庞加莱在牛顿力学路途上插上“此路不通”的牌子，热力学的创始人没有加以理会，他们继续苦干自己选定的工作。他们声明过的目的就是从原子分子方面表达世界的行为。玻耳兹曼和吉布斯与庞加莱毫无关系地设计了一个漂亮的方法来强调不同的两种系统，一种是行为可预测的简单系统，另一种是具有众多分子、与日常物体更为接近的复杂系统。

为了想象无数分子的行为，他们把这行为画在一个“相空间”里面。这绘画式方法在此学科整个专门讨论中扮演了主要角色。相空间

中的行为肖像，信息量极为丰富，它们对其中运动的显示就和从笔触辨识画家一样可靠。它们也有助于揭示时间之箭。

对在一个盒子里运动的一个台球，要画它相空间肖像，我们必须说它的位置是什么，速度是什么。位置用3个坐标表示，习惯上用x、y、z代表左右、上下、前后3个方向，速度也用这3个相互垂直方向的分量表示。这样一共要6个坐标来画这台球的行为；2个台球，就要12个坐标；3个球，18个坐标；等等。N个球，要6N个位置速度坐标：这6N维度组成我们相空间肖像的布局。这N个球在某个时刻的状态就由这6N维的相空间的一个点来代表。虽然不可能想象任何大于三维的空间，我们至少直觉上可以理解，100万个分子，用一个600万维相空间中的一个点表示，比用平常三维空间中100万个运动的点，更为方便。这些球按照牛顿的方程跳来跳去，那代表点在相空间就描出一个相应的轨道。

我们现在考虑一个完美的、来回摆动不已的钟摆的相空间肖像。每次来回，钟摆对其摆动终点的势能和它速度达到最大时的动能进行交换。如果摆动幅度不大，钟摆就是个可积分系统，相应的牛顿方程就能有精密解。这时候钟摆的相空间肖像是一个点，沿着一个环永远不停地循环[图30（a）]。每循环一次相当于钟摆的一个摆动周期。如果我们把这理想化系统拍成电影，然后既顺放又倒放这电影，那就无法判断哪个是肖像实际进行的方向——两种运动都是允许的。这里时间箭头很明显地被遗留在外面了。而这里的轨道是个闭合的环，这事实鲜明地说明了庞加莱的循环时间和永远的回归。用我们的比喻来说，画这幅肖像的画家可以说是来自可预测、可积分系统画派。

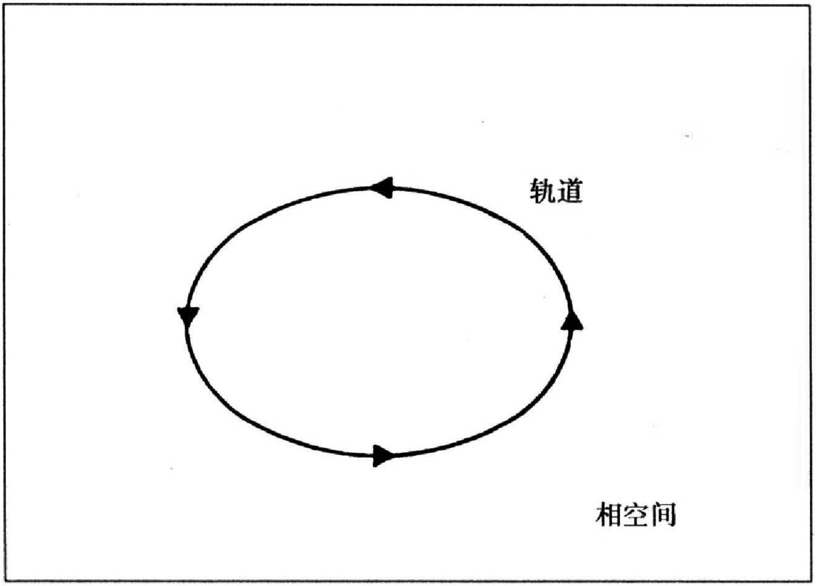
现在我们把这幅肖像和描绘一个气体分子行为更复杂的肖像加以比较。这里上亿的分子不像钟锤受着牵连，并且它们是不断在相互碰撞，在和容器的墙壁碰撞。每碰撞一次，气体的相空间坐标就做快速而重要的改变。玻耳兹曼和吉布斯经过推究，认为如果分子足够多，时间足够长，整个的相空间每一点都会经过。直观上我们可以合理地

预期，一个单独的分子，随着时间的流逝，将会造访容器中每一点。他们把这性质叫做“遍历性”。一个长生不老的遍历性猴子，在一架打字机上随意乱打，经过长远到难以想象的时间以后，将会打出莎士比亚全集（以及狄更斯全集，乃至所有其他人的所有作品）。一个像孤立的、盛有气体的罐子那样的系统如果是个遍历性系统，气体分子就会探讨相空间所有准许的地方[图30（b）]。和钟摆比起来，这里飘忽不定的轨道显示着气体分子的随机性运动。

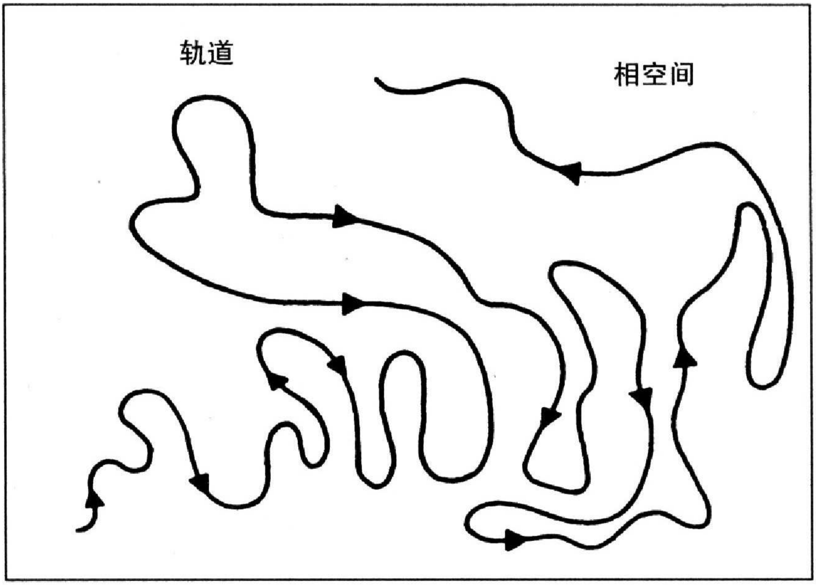
玻耳兹曼、吉布斯和麦克斯韦认为气体行为只受一个约束——能量的约束（他们考虑的是孤立系统，所以能量应该守恒）：“为了直接证明热力学平衡问题，唯一需要的假设就是：如果对在某个实际运动状态之下的系统，我们听其自然，它迟早将会经历每个与能量方程相容的相点。”难道这些遍历性系统含有了解时间之箭的秘诀？

钟摆是一个可积分系统，一群分子被假设是个遍历性的系统，它们迥然不同的相空间肖像是区别这两种不同行为极有价值的直观教具。

在20世纪30年代，数学家诸如纽曼、贝克霍夫、霍普夫、哈尔末斯（P.R.Halmos）等，为遍历性系统的理论处理，建立了一个数学上非常严格的框架，叫遍历理论。该理论此后发展成为一个完整的纯数学学科。一群（前）苏联数学家，最初经由辛钦（Aleksandr Khinchine）的启发，后来在柯尔莫哥洛夫（Andrei Kolmogorov）、阿诺索夫（D.V.Anosov）、阿尔诺耳德（Vladimir Arnold）、西奈伊（Yasha Sinai）等人的领导下，逐渐把持了这领域。




(a)




(b)

图30　(a)小振幅钟摆的相空间肖像。这是一个可积分的动力学系统。其轨道被限制在相空间很小的一个区域。（b）一个气体中分子群体的相空间肖像。这里的轨道将经过相空间的每一个部分——运动是遍历性的。

他们的工作揭示了，遍历系统中存在有不同层次的行为——有的简单，有的复杂，有的很奇妙：简单同时又复杂，它们不同的相空间肖像需要用整个一个画廊来装。就如艺术史专家把绘画分成“经典派”，“印象派”，“立体派”，“现代派”等等，这些肖像的分类可以揭露其中动力学不稳定性的特征——这和混沌有密切关系，对时间箭头在原子分子层次上的了解极为重要。

有个问题必须解决，我们才能找到时间箭头——因为即使在遍历性系统中，庞加莱回归的阴魂仍然不散。一个系统如果被锁在永远回归的循环之中，它就不可能有时间箭头。正如庞加莱所指出，一个动力学系统，不管多么复杂，如果注定它行为重复，就不会存在有伴随不可逆性的、永不回头的熵增。把庞加莱回归和无时间宇宙割开的是两个字：混沌。

## 宇宙尽头的电子

报道一场板球比赛的新闻记者，打电话到他办公室说：“史密斯把球打过外场员，赢得三分。”（Smith hit the ball over the slips and into the deep for three.）第二天，报纸上登的是：“史密斯把球打过悬岩，掉进蓝色大海。”（Smith hit the ball over the cliffs and into the deep blue sea.）同一体裁更熟悉的例子是指挥官下的命令：“派后援来，我们要进攻了。”（Send re info rce ment, we're going to advance.）这命令经过战壕里的口传，最后变成：“寄三毛五来，我们要去舞会了。”（Send three-and-four pence, we're going to adance.）

这两个笑话告诉我们，一句话稍微变一变，意思就完全两样。动力学混沌与此类似：初始条件稍微变一点儿，时间上的演化就迥然不同。动力学混沌这词有别于随机性混沌，后者是来自外界影响的真正偶然性的行为，本章不再予以讨论（这两种混沌的区别，第六、第七

章已详叙过）。在能量可以损失的耗散系统的情况之下，我们已经提到过洛伦兹用以解释天气变化无常的蝴蝶效应。它强调我们现在离开拉普拉斯式决定论的梦想世界是大有一段距离了。

“宇宙尽头的电子”是洛伦兹蝴蝶效应的宇宙版本（但这里应用到一个能量不损失的守恒经典动力系统），是英国布里斯托尔大学的贝瑞（Mi chae l B erry）将它普及的。设想我们要追踪在屋中一个很小区域疯狂运动的一个氧分子。要这样做非常不简单，因为这个氧分子要遇到成万上亿别的分子，要和它们碰撞。碰撞使它转向，每次转的角度，原则上可以算出。为简单起见，我们假设分子是像台球一样地行动。但是对这分子初始位置速度的知识，只要有一点点微小的不确定，就会很快导致对偏转角度很大的不确定。这分子也许就碰不到本来应该碰到的另一分子，这样就大大地改变了它的轨道。假设我们像上帝，精确地知道它的初始条件。即使在这种情况，如果仅仅忽略小如这分子和位于宇宙“边缘”的一个电子之间的引力作用，我们也就再没有希望预言那分子的去向了。

我们并不需要考虑宇宙边缘的电子，我们甚至连实验室几尺外的电子也都不必管，除非我们在做极其敏感的实验。想预测这个氧分子的古怪行径，不管这些，已经够麻烦的了。贝瑞的故事之有效力，在于他假定在某个特定时刻我们已经精确地知道那个氧分子的状态。可是我们将要强调，即使在牛顿动力学，它的状态也不会被知道；和此内禀的不确定相比，宇宙尽头的电子的影响是微不足道的。

## 混沌粉碎了决定论

混沌存在的场合对初始条件是极端敏感的。牛顿方程能作短期预言，不能作长期预言，除非初始条件精确地知道。真实世界的这个特点，便被用来作为同时打发决定论和庞加莱回归论的武器。在上面的例子中，气体初始条件一个小小的改变就意味着本来要碰撞的分子不

再碰撞，本来不会碰撞的分子却要碰撞。相空间代表所有分子的代表点，其轨道就要有很快的变化。这样，虽然牛顿方程应该能描述这气体的行为，但是要从这决定性运动方程中对未来做准确的预言，初始条件（所有亿兆个数值）必须以无穷高的精度得知才行。这工作即使原则上能做，也不是任何人脑、任何容量小于无穷的计算机能做的。说精度追求将是无穷，就是说这工作一直做下去，永远做不完。只有进入宗教的领域，决定论才存在。千真万确，事实上，也只有全能如上帝者，才能处理无穷多的信息。

气体是非稳定动力系统的一个例子：初始条件中微小的变化导致它长期行为中巨大的变化。类似地，不管弹球迷怎么想尽方法把钢球打得跟上次一样，结果总是不同。一个动力系统，如果是高度的不稳定，它便叫做混沌式：在此情况，初始条件很靠近的两个轨道很快地（指数式）分离。然而描述时间上演化这问题，还不仅是搜集必须放入牛顿运动方程的数据这项实际困难。假设射人弹球台的弹球的初始速度——初始条件之一，是一个介于0和1之间的数（也可以是介于1和100，或者两个任意数之间，道理一样）。这听上去很平常，其实却蕴藏着一个基本问题。因为我们用来描述这初始条件的任何数都是特殊的，都是非典型的。这是因为我们只能用所谓的有理数，它们由两个整数的商定义，可是数学揭示了众多的无理数，它们要讨厌、麻烦得多，描述一个无理数得用一个无穷长度的随机数字系列。对决定论的打击是由于，虽然在O与1之间有无穷多个有理数，无理数比有理数还更多，更无穷地多。我们只能处理有理数，而无理数必须用有理数近似。这样，我们能处理的数其实是一个极不正常的选择。当钢球开始运动时，它的速度是个无理数的概率远远（无穷远）地大于是有理数的概率。我们永远不能精密地描述它在弹球台的行动。这是原则问题，不仅是实用问题。即使有理数也会很长，甚至需要无穷个数字来表达。例如1/3是0.3333333……一直到无穷，作数字计算时必须截断，好比说，0.333。可是任何截断将会很快地导致与用“精确”初始条件完全两样的钢球轨道，产生完全不同的一场游戏。初始条件多

少总是不确定的，这事实我们躲避不了，必须正面对付。

只是最近人们才注意到，问世已3个世纪的牛顿运动方程所描述的经典系统是不稳定的，也才意识到牛顿式决定论是有缺点的。一度是传统学科的流体力学的权威，赖特希尔爵士（Sir James Lighthill），最近代表几世纪来梦想实现决定论的众多科学家，做了一个感人的公开忏悔："今天我们都深刻地感到，我们的前辈对牛顿力学惊人成就的崇拜，使它们在可预言性这领域中做了些推广，这些推广我们在1960年以前都倾向于认可，但现在我们知道是错误的。我们以前曾向知识界宣传过，满足牛顿运动方程的系统是决定性的，这在1960年后的今天，已被证明为不正确。我们在此集体向知识界道歉。"赖特希尔本人的专业，从前是工程师和应用数学家独占的地盘，现在经由对动力系统的新处理方法，已成为数学物理中新颖而硕果累累的一门学科。

无数考卷叫学生把牛顿方程应用在圆球的碰撞或行星整齐的轨道运行上，要知道这些问题都是例外，不是普遍现象。认为"正规"教育是和"现实"脱节的人，大可以用此作为武器。虽然用来处理这些问题的方程很简单，混沌这概念的妙处就在于，从同一个源，也能产生复杂的行为。

经过遍历理论的发展，相空间中的复杂性和时间箭头变得容易想象（图30）。混沌的手法可由相空间肖像显示。我们上面看到，初始条件，即使对一场弹球游戏来说，一般也是不能精密地确定。考虑到此点，相空间肖像不应该再用单独一点，而应该用个一小团。这小团包含所有与初始条件不确定性相容的轨道（气体分子的也好，弹球游戏中钢球的也好）：这小团包含可能性的范围。这小团在时间上将如何运动？游戏规则很简单：小团按照刘维方程演化（刘维方程，第五章讲非平衡统计力学时已讨论过），小团的体积必须守恒，形状不一定守恒。体积守恒的理由可以用容器中气体来说明。不管气体有多少种分配给诸多分子的方式，气体处于容器中的概率总是一样，总是等

于一。气体总不会不翼而飞。这样，这小团可以看为一滴不可压缩液体，永远保持体积不变——因此保持总能量不变，但可以变化形状（相空间中整体形状标志运动是如何分配在分子中的）。

现在让我们来浏览一下相空间肖像画廊（这些肖像都不是用“一点”而是用“一小团”画的），看看是否能看出不同画家的手法（图31）。图31（a）是主考老师的得意之作，简单美感的曲线说明这是个可积分（非遍历性）的系统，这里的小团做周期性的旋转，只经历整个相空间（“布局”）的一小部分，并且保持它的形状。此画派肯定也影响了图31（b），那是一个遍历性系统，处于小团中所有的轨道仍然始终彼此靠近（不然小团就变模糊了），但是小团经历整个相空间。肖像（a）和肖像（b）代表稳定的（或规则的，或非混沌的）动态，因为初始条件中微小的不确定性，在长时间以后的系统状态中导致的不确定性，仍是同样地微小。

肖像（c）揭示的是可能更有兴趣的情况。它的狂舞，像一幅贾克逊·泊罗克（Jackson Pollock）的画：这运动经历整个的相空间，因此是遍历性的。可是小团的体积总是不变，它的形状却变成越来越长、越来越细的细丝，变成像掉进水里的一滴墨汁，放进咖啡里的一团奶油。到了它最后渗入布局的每个部分之后，它就不再演化了。这点具有重要意义，因为它告诉我们，在概率分布函数（小团）的层次上，可以达到平衡，时间演化可以有个终点。因为这种肖像和不同液体的扩散混合颇有相似之处，所以它叫做混合式遍历流。霍普夫第一个研究了这种时间演化的数学细节，虽然这种行为吉布斯已经想到过。

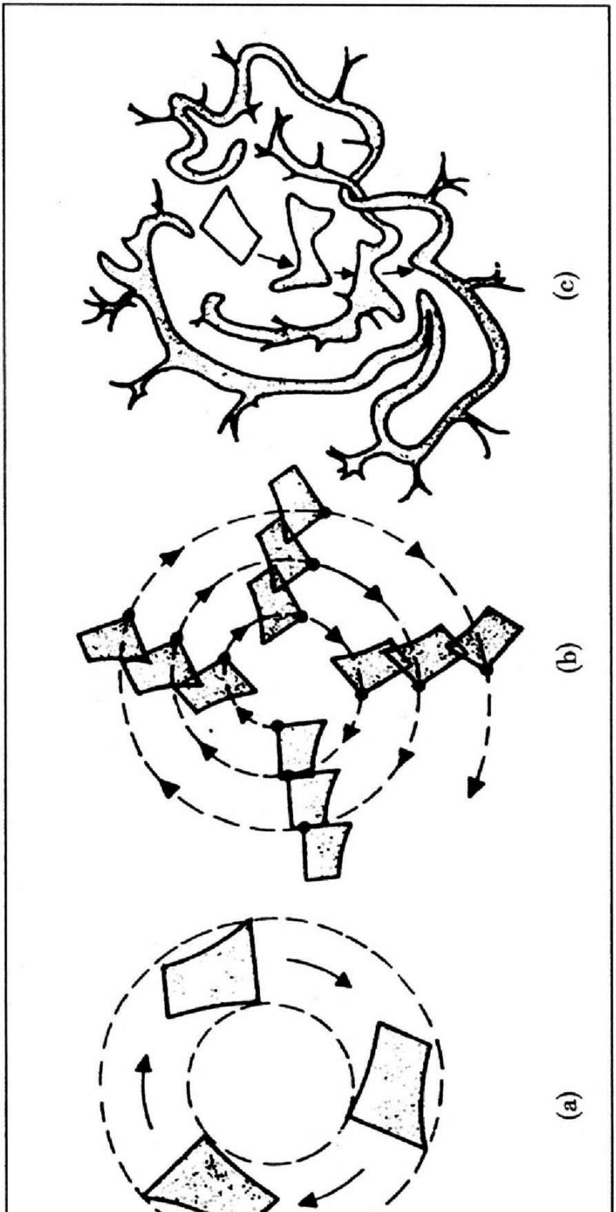



图31　相空间概率密度的时间演化；（a）非遍历性；（b）遍历性；（c）混合式。[录自巴力斯古《平衡统计力学与非平衡统计力学》，第718页。]

混沌的作用可以在图32中看到。这里显示的是肖像中“笔法”的细节，在初始条件小团中邻近两点之间距离随时间的变化。

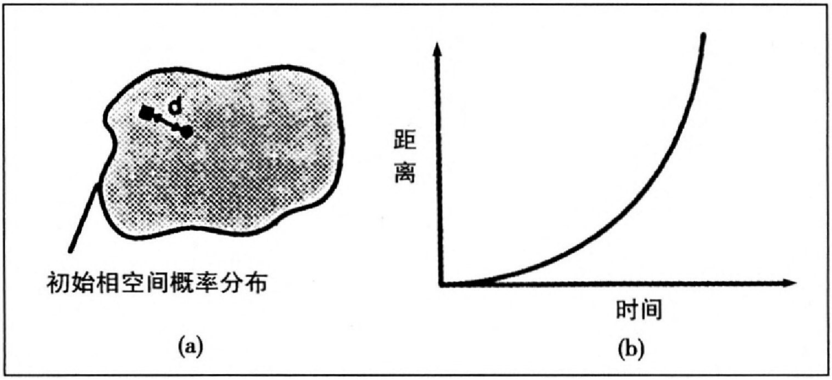



图32　混合式流动的动力学不稳定性。（a）初始相空间概率分布中相聚d的两点。不管d是多小，这两点将随时间指数式地分离，如（b）所示。比较图31（c）。[录自柯文尼，《研究》第20卷，第190页（1989）。]

不管两点最初靠得多么近，它们总随时间指数式地分离。在很短时期内，这样两点的行为差别不大。但过了一段时间以后，它们的长期轨道就完全不同，它们各自经历相空间中完全不同的区域。这正是我们所谓混沌式时间演化的意义，因为只有我们能无穷精确地得知初始条件，我们才能利用牛顿的决定性方程来计算未来的行为。用相空间绘画语言来说，如果开始是单一的一个点，我们就可以求助于牛顿。但是对混沌系统，初始时刻不会没有的不确定性意味着牛顿物理的基石——可预言性——瓦解粉碎了。

对于这种混合型流动，一丝不苟的决定性必须退位给概率陈述。这就是说，我们必须一开始就放弃决定性的轨道，而只运用概率——这正是统计力学的办法。此结论对时间箭头颇具意义，科学家们，尤其是玻耳兹曼，想用分子运动来说明不可逆的热力学第二定律，既然他们用的主要方法之一就是统计力学，因此我们刚得到的结论，对时间箭头颇有意义。要注意：我们达到这结论，并没有引用任何主观论

证。

这混合型肖像只是整个混沌肖像“族”里的成员之一。如果把遍历性系统按照不稳定程度即混沌程度排列，混合型流动是居中的。比它更混沌的行为发生在所谓K型流动中。这里，K代表柯尔莫哥洛夫，他和西奈伊研究了这种流动的性质。K型流动的行为处于完全不可预言性的极端，尽管“内中的”运动方程仍然是决定性的。K型流动具有如下令人注目的性质：初始即使有无穷多个测量，也不能预言下次测量的结果——除非初始测量是无限准确的，物理上那当然不可能。这种流动本质上是随机性的。

所有这几种行为可能听起来都有点太抽象。遍历性系统跟实际世界之间究竟有多少共同之处？事实上，直到西奈伊的创始性T作出现以前，很多人已经开始怀疑，现实与嵌在遍历性理论中抽象数学之间，是否有任何关系。但1962年，西奈伊宣布，他证明了一个只盛有两个或多个按照牛顿方程运动的台球的盒子，也具有图31（c）所示的混合性质。西奈伊的结果是对决定论又一个打击。仅仅两个球的运动，虽然比较理想化，确也具有有统计力学所研究的行为的性质。而直到那时为止，人们普遍以为，遍历性（更不用说更强的混合性了）只是巨大数目的原子或分子（例如一个气体中成亿成兆的分子）才有的性质。西奈伊指出，如果盒子里的球不止两个，那么动态就更退化一步，成为K型流动。这样，即使一场台球游戏其实也是混沌式的，也是不可预测的。球杆碰球的情况，稍微有点变动，球的长期位置就完全不确定了。幸亏对台球球迷来说，这不可预测性只是在击球完毕很长时间以后才能察觉到，并且是在不存在摩擦力的情况之下才行。

西奈伊创始性工作以后，许多别的理想化的情况，其中有的涉及台球彼此之间碰撞和台球与凸面边界碰撞，也都被证明具有混合式流动甚至K型流动的性质。从理论观点看来，缺点是建立遍历性性质极其困难，一个典型的严格证明长达数十页。已证明具有遍历性性质的情况都比较“奇异”，例如台球式的相互作用不能作为分子作用的典

型，因为台球直到碰撞以前，根本不管彼此的存在。真实世界里的相互作用总是更平滑些。不过，很多人认为这只是一个技术性的困难，认为真实系统大多可以用遍历性的K型流动来模拟。

## 混沌与时间之箭

对于相空间肖像画廊如何跟时间配合这问题，现在让我们来试图给出一个更坚实的理解。我们已经看到，一个混沌式混合型系统（例如盒子里的气体），它的概率分布开始是一个小团，然后逐渐生出越来越细的卷须。这种卷须的成长和探触标志着气体分子对多得令人头晕的可能性进行调查。用单独一点的一定轨道来代表气体所有分子的运动是无意义的。我们唯一能说的是：气体分子，开始如果是在相空间某个小体积中某一点，现在就可能处于肖像中某条卷须中的任何一点。这里存在有巨大的可选择性。这也许就是微观层次的时间箭头，因为小团具有一个毫不留情的倾向，把自己变得越来越模糊，最后达到平衡的死亡状态，那时整个相空间布满细丝——变成了一团概率均匀的棉絮。

我们一旦放弃基于轨道的决定性描述，我们其实就已经根据对甚至是最简单的情况的了解，彻底进行了理论上、哲学上的再估价。我们不再有一个僵硬的决定性框架和固定的预言本领，我们被“降级”到统计层次，那里不存在决定性，那里的未来行为只是在一种概率意义上可预测。

以前没有不可逆性的场合，现在可以为不可逆性腾出地方来了，并且做法完全是客观的。它和动力学不稳定性密切相关，而这不稳定性是系统中的一个内禀性质，一个客观性质。因此即使在只有少数几个台球或分子的情况下，也运用概率并且承认这是基本做法是有道理的。这做法是基于系统的内禀性质，而和我们的干预、粗粒化方法以及其他“主观”行动无关。

以前认为是不允许的东西，经过运用概率理论对非稳定遍历性系统的描述，变为可能了——其中包括不可逆性，从而包括时间之箭。可是庞加莱的回归论不再来捣乱了吗？它不是说谁也逃不过永恒的回归吗？量子芝诺佯谬（见第五章）发明者米斯拉证明了：当我们放弃用个别点子画成的肖像，而采用基于小团的肖像，用基于概率的图像代替基于轨道的图像，庞加莱回归就不再成立。

关键是：只有采用概率途径，而不是用牛顿的决定性方程，我们才可以合法地寻找一个量，它类似于熵，从它能得到第二定律的时间箭头。米斯拉在1978年的一篇重要论文里，证明了对于K型流动可以找到这样一种熵。上面我们看到，K型流动包括台球和理想气体，一般认为在大自然中很是普遍（这点尚未被证明），尤其是牵涉到宏观系统的场合。就像热力学的熵一样，米斯拉的熵也是在平衡态时达到最大，那时概率分布停止演化。

借助动力学非平衡熵，我们把牛顿学说的盖子揭开了，发现它里面的情况其实是和平衡这热力学概念相容的。这对我们在微观层次上寻找时间方向是一个紧要的发现。这是因为，就如第二定律所示，平衡状态其实就是时间之箭的箭靶。至于这箭头指向“前”，还是指向“后”，是另一个问题。回忆一下，作为此分析基础的经典力学是时间对称的：理论上可能有两个不同的时间演化，一是朝着未来的热力学平衡，一是朝着过去的平衡。

这两种演化应该选哪个，这问题对我们现有的基本原则来说太深奥了，是无法解决的。答案很可能归根到宇宙学上去。也许根据彭罗斯的建议，我们在第五章考虑宇宙学时间箭头时曾讨论过这建议，不过这还是属于臆测。我们采取的做法是：把第二定律当做自然界一个唯象事实，用它根据观测来选择朝向未来的演化。我们采取这简单而意义重大的步骤以后，就能把第二定律的时间箭头加进动力学的结构之中，从而使后者彻底变质。在此解决不可逆性佯谬的办法之中，第二定律和力学的地位相等，这办法和简并派不成功的方法迥然不同，

历史悠久的简并派总想用力学来说明热力学。

这办法从米斯拉后来和普里高津、库尔巴伊（Maurice Courbage）的合作中得到更多的支持。他们证明了如果存在一个类熵量，那么可逆性的轨道便不能用。他们并且还发现了与不可逆性相容的时间模式，叫做“内在时间”，它标志着一个动力系统的年龄。这年龄反映着系统的热力学情况，而牛顿方程中的描述则表达纯粹动力学可逆性的特征。

自从不可逆性佯谬问世以来，人们100多年来把力学和热力学放在敌对位置。而布鲁塞尔小组在两者之间建立了一个令人注目的关联。海森伯的不确定性原理（见第四章）所表示的、玻尔对量子力学的解释所强调的量子力学中观测量的互不相容，我们在这里发现一个和它很类似的情况：一个系统的热力学性质（即系统的不可逆年龄）如果完全知道，可逆性动力学描述就失去意义；反过来，动力学描述如果完全确定，热力学观点就立不住脚。对此情形，普里高津做如下说明：“世界之丰富多彩，不是一种语言可以尽括的。音乐不能完全归纳在从古典作曲家巴哈到近代作曲家希恩堡的各种风格之中。我们经验的各种不同方面，同样地不能用单一的描述一言而尽。”

分别以动力学和热力学描述的可逆性和不可逆性，看上去是同一个硬币的两面。就如我们在量子力学中发现的情况一样，世界整个结构太丰富了，我们的语言说不完，我们的头脑懂不全。对时间这新的看法具有两个紧要而互补的方面，它们之间对比之强烈，犹如天堂和地狱。统治天堂的是动力学方程，它们是可逆的，是“无时间”的，它们的单纯保证着永远无穷的稳定。地狱则像是实际世界的近亲，其统治者是起伏、不定、混沌的。这是个不稳定的，朝着死亡与平衡衰退的世界。

## KAM来到

这样，在经典物理学的核心中发现了能描述年龄概念的一种新时间；它和随时间增大的一种熵联系着。这样一来，不可逆性和熵看来都是充分不稳定的动力系统的基本性质，即使系统只含有两三个相互碰撞的物体。我们难道在热力学时间箭头与统治微观世界的可逆性方程之间，得出了一个普遍的联系吗？还没有完全这样：别忘记我们还没有考虑相空间肖像画廊中所有各种运动，我们只考虑了可积分系统和遍历性系统两种。要知道此外还有牵涉到三个或更多个物体的情况，这些都是不可积分的，这是庞加莱首先发现的，本章上面已经提过。

我们处理过的相空间肖像是两种极端情况——极端“简单”与极端“复杂”的。遍历性系统，特别是混合型流动或K型流动，属于“复杂”系统，因为即使是两个或三个台球在一个盒子中的运动，相空间肖像是毛发状的，表示整个的相空间都将被经历。另一方面，一个理想化的钟摆或者一颗绕日运行的行星是简单（可积分的）系统的例子，意即它们的运动受着高度限制，没有碰撞，可预测，因而是非遍历性的；它们的行动单调无味，有规有矩。这些可积分系统的相空间肖像是由受限制的（非遍历性）轨道组成，这些轨道周期性地重复咬自己的尾巴。这定性上和我们自己对简单系统的经验相符，“咬自己的尾巴”相当于地球重新开始一个绕太阳的循环，钟摆下一次的摆动。

这看来颇令人不解，尽管一个稳定的地球轨道让我们非常放心。从我们上面得到的认识，地球的绕日运动似乎具有一个随便乱来的混沌系统所具有的因素：它应该不是一个可积分的“二体问题”，因为还有月球和其他行星的引力作用。天体的运动其实是多体作用的结果（天体其实不止两个），因此，按照庞加莱的理论，这运动应该是不可积分的。而我们不是已经看到了除掉最简单的情况以外总应该出现混沌？

幸好我们仍可指望太阳每天升起，亏的是一个以柯尔莫哥洛夫-阿

尔诺耳德-摩塞（KAM）命名的定理。此定理出于柯尔莫哥洛夫1954年开始的关于庞加莱回归的工作，这工作随后在20世纪60年代初期由他的同事阿尔诺耳德加以发展。摩塞（Jrgen Moser）同时先后在德国和美国用类似的方法做了同一工作。他们的研究证明了庞加莱的这种系统属于完全规则和完全混沌之间的一种中介情况。

这种行为半明半暗的区域，可以从在简单情况里加一剂新配料以后达到。在我们地球绕日运动的模型中，这新配料就是我们引入的一个或多个代表月球或其他行星的小小的力。人们起先以为这种摄动肯定会把非遍历性转化为遍历性，把一个简单的相空间肖像，例如描述钟摆的环，改变成一个复杂的肖像，例如我们已经遇到过的“毛茸茸”的混合型流动。

可是KAM定理表明，这不一定发生：这小摄动在布局的不同部分，作用不同。布局有些部分仍旧是规则的，就和未加摄动以前一样，有些部分就出现了很不规则的、混沌的行为。前者的初始条件小团仍旧被限制在比较简单的闭合环体里面，而后者的小团则伸出细长的卷。图33示意一个简化过的例子。规则区域相当于稳定的运动，相当于物体之间没有碰撞；混沌区域相当于物体经受了碰撞的随机化作用。看来好像这幅肖像是两个画家画的，每个画家画了不同的部分。这两部分一般很复杂地彼此缠绕着。

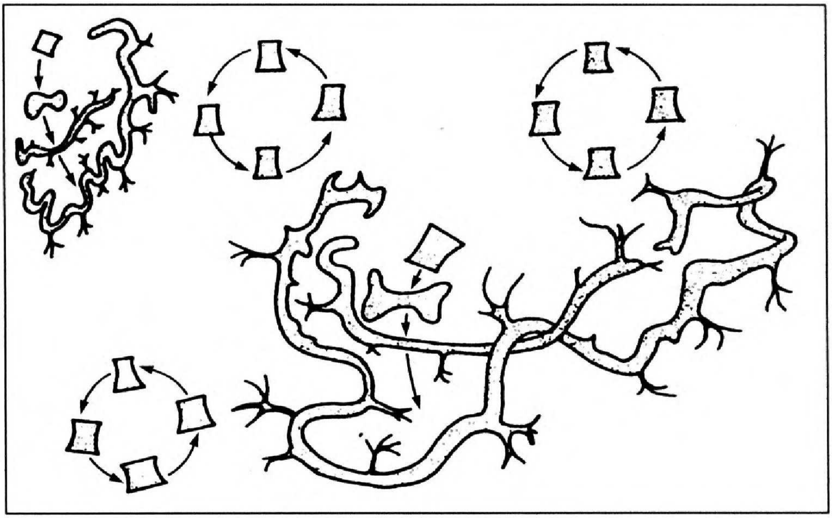


图33　一个不可积分系统相空间行为的简化图。规则行为区域和无规则行为区域同时存在。参见图31。

许多不同的因素支配着布局应该如何分成规则和无规则的两种区域。假设开始的摄动很小。我们逐渐增大输入系统的能量，我们将观测到一个所谓“随机跃迁”：本来主要是很规则的运动变成主要是很不规则的运动。原因是：能量越大，碰撞越有机会破坏和平的规则运动。此外，外加摄动的增大如果超过某个值，所有的动力学运动，由于同一缘故，都要变成混沌式运动。

天体力学界的天体物理学家听到KAM定理的消息以后都非常兴奋，因为它为证明行星运动是稳定的这尚未解决的问题大大向前走了一步。行星运动如果是不稳定的，那么像地球这样一个行星就会一声不响地掉入深渊。生命也许从来就没有出现过。即使出现了，天空中的混沌行为也不会给古人什么宇宙“内在规律性”的线索了。在第二章我们看到了，这种鼓励对发展智力思考（以及可逆性时间）曾起过极大作用，没有这种鼓励，牛顿科学本身也许不会升到统治地位。

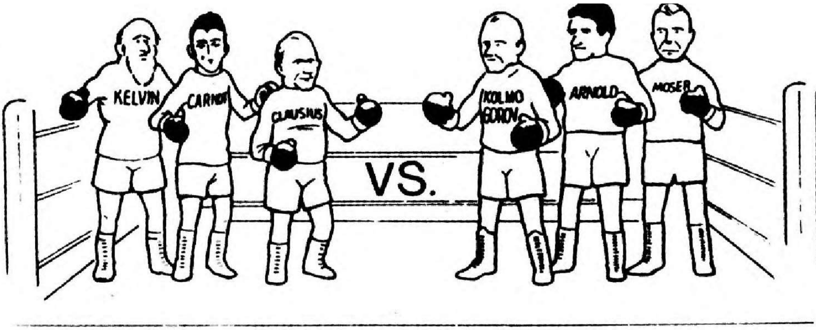



图34　KAM定律对热力学第二定理的挑战。[录自《新物理》（P.C.W.Davies编辑）中福特的论文的图12.14。]

就如KAM定理所预言，在相空间，某些部分运动是无规的，在太阳系中我们的确能找到一些粒子。状若土豆的土星第七卫星，它的运动就是混沌式运动。它围绕土星的轨道大致是规则的，但它同时不规则地、不可预测地打筋斗，就像一个土豆在地上打滚一样。冥王星的轨道中也有类似的行为。火星与木星之间的小行星带中的空隙，混沌动力学也给了解释。

对于日夜在力学中寻找时间箭头的人，KAM定理乍看上去不是好事。他们才刚以为终于能把时间的描述放在动力学不稳定性和随机性的基础上，KAM定理又来捣乱，说复杂系统能在相空间某些部分呈现简单、无时间性的行为，和时间之箭能在相空间另外一些部分出现一样地肯定。然而，重要的计算机模拟工作（特别是亚特兰大的乔治亚工业学院的福特与其合作者所做的）表明，在“物体数目充分大”的情况下——具有成亿成兆分子的宏观层次是肯定满足这条件的——所有关于规则周期运动的行迹都会被淹没掉。那时，KAM定理将失去作用，遍历性、不稳定性、从而不可逆性再次主宰一切。虽然我们应该说明，这些话还没有给予严格的数学证明，时间之箭看来经过计算机的数字啃嚼而复活。这和我们自己的经验相符：正是在日常物体、事件（例如在融化的雪人）的层次上，我们最感觉到时间的箭头。

可是，这里看来好像也有一个截止点和粗粒化中的差不多。第五章介绍了那里的截止点，它说当我们从原子尺度移到苹果尺度的过程中，时间之箭会神秘地出现，因此我们将粗粒化轻率地打发掉。现在当我们从少许几个原子移到“数目充分大”的情况的过程中，时间箭头的出现是否也具有某些随意性？统计力学启示我们利用“热力学极限”办法来处理这问题。我们不要设想一个任意大的分子集合，让我们考虑如下的系统：它的分子数目N和它的体积V都趋向无穷，但分子的密度（N除以V）一直保持为有限。热力学极限避免了时间箭头在某个随意尺度上的出现。这是统计力学不可缺少的一手，应用很广，例如固体转化为液体时熔点温度的理论计算。在热力学极限下，KAM定理失效；任何“无时间性’的规则行为都被冲走，只有无规则的混沌式运动留下来。这样，时间之箭就显露出来了。

## 相对论混沌

关于相对论力学中的混沌只有少许认识。第三章提到，广义相对论中，运动是沿着两点之间距离最短的测地线进行的。法国数学家哈达马（Jacques　Hadamard）在19世纪末证明了，在一个曲率为负常数的面上，测地运动是高度不稳定的。此后又证明这运动是一个遍历性K型流动。在某些宇宙学场合，可以证明这种测地运动是的确发生的，于是在这些场合中可以建立内在时间和年龄的观念。此外，甚至相对论不变场——用以描述电磁现象的那种，看来也有K型流动的性质，因此也具有内禀的不可逆性。可是在有引力相互作用的场合中，证明诸如动力学熵、动力学年龄的存在，殊非易事，至今尚无答案。

## 量子论与不可逆性

至此为止，我们所考虑的是：当我们用牛顿的观点来处理微观层次上的事件时，不可逆性会以何种形式登上舞台。可是近代的证据（第四章讨论过）是说：要正确地描述微观物质，必须运用量子力学

的语言，尽管这样做有许多困难。因此，我们应该设法把不可逆性的讨论建筑在量子理论上。这样做和用经典理论有许多相似之处。

在量子力学中，一个系统的状态由波函数描述，而对于嵌置其中的可观测量（诸如位置、速度），测不准原理保证它们总有些内在的不准确性。然而，被薛定谔方程决定的波函数的演化是可逆的，就像牛顿力学中的轨道一样。因此，很自然的一个问题是：是否存在一种波函数的动力学不稳定性与经典力学中的混沌类似。这样一来，我们必须把第五章提到的量子力学密度矩阵拿来取代波函数，作为考虑的基本对象，就如在动力学不稳定的系统的经典力学中，几率分布函数是它的基本对象一样。

尽管多年来许多人的认真努力，在小的量子系统中，我们（至今）还没有找到与混沌类似的现象。这和量子理论中存在一种强形式的庞加莱回归有关，它说：所有有限的、隔离的量子系统。比如放置薛定谔的猫的装置都是周期性的——它们都要永恒地回返。薛定谔的猫将永远陷在既活又死的暧昧状态之中。这样，量子力学中似乎没有一条趋向平衡的单向通途。然而，只要承认存在具有巨大数目分子的宏观物体，就可以把一种熵引入量子力学。办法是采用热力学极限，就像经典力学对大的、不可积分的庞加莱系统的处理一样。当然，和上面那种情形相同，只是在考虑宏观系统而不是考虑微观系统的时候，不可逆性才可以这样出现。

对于量子力学的主要困难之一——第四章讨论的测量问题，这一结论颇为重要。虽然量子力学的实际应用已成惯例，但是对它应该如何理解仍有可争论之处。为了事件在真实世界中发生，波函数在测量时必须坍缩，这时信息是如何从理论中提取出来，这便是最大的难处。量子力学在微观层次上是非常成功的，那么现在大惊小怪些什么？

贝尔（John　Bell）是欧洲粒子物理中心（CERN）的一位理论物

理学家，他对量子力学的内在问题做过深刻的思考。他采取的是“悄悄地”态度。他认为：“量子力学里基本性的含糊并没能阻止我们进步。我们的理论学者在那含糊中迈步前进……已取得的成果是极为动人的。如说这是患梦游病的人做的，我们难道要把他叫醒吗？我觉得是不叫为宜。因此我现在把声音放低，悄悄地说话了。”

如果一个人喜欢在微观世界中生活，那也就不妨梦游梦游。可是从第四章得出的结论是：量子力学用来描述我们自己的世界时，发现它在宏观层次上是有问题的。困难可以用像薛定谔的猫那样的量子佯谬来说明，其中“量子实际”需得用“又死又活的猫”这类暧昧状态来描述。对此情形，贝尔概括如下：“世界究竟应该如何划分为我们可言的仪器和我们非可言的量子系统？通常的理论中的数学要求这样一个划分，但不告诉我们如何划分。”

他继续说道：“难道亿万年来，世界波函数一直在等一个单细胞生物的出现，然后才跃迁（坍缩）?还是它还得多等一会儿直到出现了一个有资格的——有博士学位的观测者？如果这个理论不是仅仅只能用在理想化的实验室过程，我们不是就得承认，差不多每个时刻，差不多每个地方，都在进行差不多的‘类测量’过程？这样一来，难道还有某个时刻是没有跃迁的、是薛定谔方程能适用的？”

一般看法认为：每做一次测量，波函数就不可逆地、完全不能预定地、随机地坍缩——因为这坍缩到底是在薛定谔方程的范围之外。可是，正如玻尔与其学生罗森菲尔德（Leon Rosenfeld）一再强调过的，测量这过程是用一个宏观的仪器，在宏观世界与微观世界的“交界处”进行的。按照上述办法把熵和不可逆性引入量子理论框架，对于这个疑难问题，可以得到一些认识。如果可以找到一个类熵的量，可逆的波函数就可以被一个概率手段所取代，就像牛顿方程被刘维方程所取代一样。不可逆性一旦被承认，波函数坍缩就失去它先前那种神秘性。从这个观点来看，测量过程根本没有什么东西令人注目——它就是与热力学第二定律相符的一个不可逆过程的典型例子。

在量子理论中引入内禀不可逆性很富有吸引力，它把热力学第二定律从开始就包含进去，并且解释了测量过程。然而，由我们看来，这仍旧是一套不完备的理论，可逆的量子定律和不可逆的热力学仍旧是以某种特殊方式连接起来的。我们同意贝尔的看法，即在这方面某些基本措施仍待发现："观察事物的新方法将牵涉到某种想象跃进，这跃进将使我们惊讶。无论如何，量子力学的描述将被取代。在这方面，它和其他人为的理论一样。但在更大的程度上，从它的内部结构中可以看到它的末日命运。它本身就载有自我毁灭的种子。"也许彭罗斯的猜想，即一个令人满意的、未来的量子引力理论中的时间是不对称的，能提供所需的"彻底的意识更新"。这样一套理论，如果成功，可以期望拿掉广义相对论中的奇点，说明热力学第二定律，解释波函数坍缩。不过这项工作属于未来。

## 时间与豹斑

这对于前几章用以描述豹子如何得到身上的花纹、黏菌如何聚合的方程，有何意义？那些含有时间箭头的非线性"运动"方程，它们的动力学来源是什么？

物理学成功地运用了这些方程——其中包括第五章提到的著名的玻耳兹曼方程来描述诸如黏滞性、扩散、热传导的运输过程。我们熟悉的不可逆过程——扩散，在它里面的是物质从高密度区流进低密度区。类似的，黏滞性来自一种液体的摩擦，由于这种摩擦，流体中的有序的机械能耗散为热能，热能相当于分子的随机运动。

可测的量，比如一个液体的黏滞性和热导率，大多数人认为是物质的"客观"性质。但是绝对坚持原子简并论的人，却认为这些日常现象是"幻觉"，应该予以摒弃。难道这些像黏滞性的性质，真是一个基本上无视时间的宇宙里面，我们目空一切的想象中的虚构？为了解释周围的世界，我们难道还得向主观因素求助？幸好我们有理由相信，

答案是否定的。这是因为现在有一个可说是普遍的方法，可以用来导出这些运动方程，这方法看来和不可逆性的来源有深奥的关系。尽管这普遍方法有不少数学上的困难，它看来能够阐明时间箭头问题。特别是：20世纪60年代晚期至70年代早期，乔治（C laud e George）与何宁（Fran oise Henin），在与普里高津密切合作之下证明了，一个大的耗散系统的时间演化，可以唯一地分成为两个完全独立的、叫做“亚动力学”的成分。

一个是“运动”成分，它描述系统的长期演化，含有到达热力学平衡态的途径。另一个是“非运动”成分，它描述初始条件开始以后不久的短暂行为，这种行为随演化而消失。我们还不太清楚，这两种行为究竟如何划分，究竟在哪个时刻分子就不记得初始条件了，就在时间之箭的影响之下奔向平衡态了。对此问题，活跃的研究仍在进行。动力学不稳定性（混沌）是一定需要的。有件事是肯定的：这种行为如能从微观动力学导出，主宰运动成分的将是一个高度普适性的、时间不对称的运动方程。这样，我们终于揭露了统计力学中的时间箭头。

亚动力学这园地，丰饶多产。它使新的和老的运动方程可以系统地导出，并来描述一大批不同的现象。巴力司古（Radu Balescu），巴黎南边丰特乃玫瑰镇CEA（欧洲原子协会）的米斯基齐（J.H.Misguich），贝尔格莱德物理研究所的西卡尔卡（Vladimir Shkarka）和本书作者之一柯文尼将此分析扩展，他们研究其演化的系统受着一个时间上有变化的外在场的支配，例如由激光产生的电磁场。这对于探求可控聚变发电是一个颇有兴趣的课题，因为后者就是把由电离原子组成的等离子体限制在一个磁场里面，希望载有正电荷的原子核彼此碰撞，从而释放大量的核能。这种等离子体的时间演化可以用非线性运动方程精确地描述，而这些方程可以用亚动力学手段得出。此外，在空间时间都变化的电磁场的情况之下，描述几种特定等离子体的演化的方程，也可以用同一手段来解。

在避免用主观的方法把不可逆的运动方程从力学中推导出来的努

力中，如果说布鲁塞尔小组是孤军奋战，那就错了。别人也试图用过另一些同样是微观的、客观的方法。加州大学伯克利分校的一位数学家兰弗德（Oscar E.Lanford III）在1975年对玻耳兹曼方程做了至今为止最严格的数学推导。可是，他只能证明他的推导只是在很短时期内有效，而我们预期它应该最适宜描述对稀薄气体的长期行为。其他一些人发展了基于“定标”技术的方法，这些人中包括慕尼黑大学的斯庖恩（Herbert Spohn）。不过这些方法有与兰弗德的方法同样的缺点。这些手段的优点是高度的数学严格性，但是它们没有布鲁塞尔小组所发展的那么多的科学内容。

## 熵与创生

不可逆性与熵看来和所有事件中最重大的事件——时间、宇宙本身的诞生无可避免地连在一起。这启动所有事件的事件——宇宙的无中生有和熵的产生不可解脱地结合在一起，因此是不可逆的。第三章讲过这个熵，它无所不在：彭齐亚斯和威尔逊在1965年发现的来自全天的微波黑体辐射，相信是大爆炸的一个遗迹。这个黑体辐射由低能光子“稀粥”组成，遍布全宇宙，它是一个丰富的熵源。

可惜的是，这个熵不可能跟描述宇宙大尺度结构的爱因斯坦相对论方程挂钩。这些方程只能描述可逆过程，因此无法解释这个熵从何而来，除非用我们已经遇到过的、定义模糊的粗粒化。如果我们因为这办法是特殊的而不加采用，那么问题仍然存在：熵是如何产生的？

均兹格（Edgar Gunzig）、戈呵纽（Geheniau）、普里高津在1987年对此问题提出的一个解答，虽属臆测，却具有魅力。这个答案建筑在另外一些宇宙学家的想法之上，其中包括，除均兹格外，有布柔特（Robert Brout）和恩格勒尔特（Fran ois Englert），他们的宇宙无中生有，在上面第四章中曾有讨论。

让我们来回想一下他们的看法：最初，时空是一无所有的，是平直的——根据爱因斯坦对引力的解释，那时没有物质使时空弯曲。海森伯的不确定性原理允许在短期间内可以免费借能量来创造宇宙。根据爱因斯坦的物质-能量关系，这能量产生物质（以黑洞形式），物质又引起时空的弯曲——也就是我们所谓的引力。从“无”，我们得到相当多的“有”；尽管如此，产生宇宙所需的总能量等于零，因为宇宙中所有引力的能量是负的，它和产生的质量的（正）能量恰好抵消。

本章前面我们试图用原子分子的语言来表达第二定律时，我们看到，混沌和动力学不稳定性这些概念与热力学不可逆性相容。这里的建议是：平直（“闵可夫斯基”）时空量子真空的不稳定性与不可预测性，经过种下充满物质的宇宙的创生的种子，导致出不可逆性。这样，物质形成的过程被认为是在宇宙学尺度上不可逆的，并且黑体辐射中的熵就是这个原始过程产生的。时间的诞生于是变为一个无可避免的单向过程。它是时间之箭的最终表现。

有人认为，宇宙创生以后，经历过一个“暴涨”膨胀阶段（其时黑洞蒸发），然后转变成我们今日熟悉的那种物质辐射混合体。这个模型中一个很有意思的特点是，我们在它里面再次看到时间的双重面目：不可逆性与重复。因为宇宙如果是开放的，也就是说，如果没有充分的物质把它拖向“大坍缩”，它就要一直膨胀下去，宇宙中的物质密度将要变为极端稀薄。这种情况相当于一个平直的时空，于是整个这场戏将要重演在大得非常非常多的尺度上。看来，我们在这本书里一再强调的时间的这双重性质，在无与伦比大的尺度上，居然也可能存在。

# 第九章　未结束的探索

我们已经达到了知识的统一，还是科学在矛盾的前提下已经土崩瓦解？这样的问题将使我们对时间的作用有一个更深入的认识。

——普里高津

《从实存到将然》

在我们试图建立时间之箭的努力中，仍然存在有许多问题。很多问题的解取决于高深的数学，然而还是为大胆的洞察和直觉留下了余地，它们可能会导致今天几乎没有想象到的新的概念。

我们已经为这不可逆佯谬苦恼了100多年：所谓的“基本”理论在区别过去和将来中遭到失败。诚然，这些可逆的理论给了我们对于世界的强大洞察力，但就是对于这同一个世界，死亡和腐烂在提醒我们，不可逆性和时间之箭有多么重要。

宇宙的诞生和膨胀也为时间之箭作证。同样作证的，还有离奇的长寿K介子的衰变；光波传播到将来而不是过去的偏向；物体混合、冷却和腐烂的倾向；以及演化树的多种多样的不对称性。只有在一个不可逆的世界里因果才能分明，因而一个合乎逻辑的对事件的陈述才可以成立。

不可逆性容许有一系列令人兴奋的可能性，因为只有有了不可逆性，才会有生命，才会有天地万物。许多人曾试图贬低它的作用；与此相反，我们现在知道如何来利用它。不可逆性可以帮助揭示我们在时间和空间中看到的许许多多变化万千的图案。它这种创造力也就是

当今一些时髦词汇。诸如“耗散结构”、“自组织”、“混沌”，“复杂性”，等等的后台。从今天的观点来看，不可能把不可逆性贬黜为虚幻一类，因为它的起因看来就是我们周围大自然中存在的动力学不稳定性。

如我们已经看到的，理解全部这种复杂性的关键，是处理非线性动力学的一个数学分支。它的应用今天已超出了“硬”科学的范围。无疑地，经济情况也一定有它的混沌方面（虽然它也同时受到随机的外部因素的驱动）。怪不得经济预报看来更像一种魔术而不像科学。

人类社会和动物社会也会形成顺着时间之箭的图像：从群体迁移到足球比赛中的万头攒动。这些社会可以看做是开放的和高度非线性的动力学系统，其中大量存在着反馈环和竞争。科学家们已经开始在寻找化学反应中的自组织和混沌和在人类及动物社会中出现的“革命”、“动乱”、“经济崩溃”等现象的对比。

不可能预测这种新的研究方法会导致什么，但是这种在物理学和社会科学之间的跨学科的交叉研究，看来对双方都有益。一方面，物理学家将了解更多的有关现实世界中的内禀复杂性；另一方面，社会科学家（甚至于政治家？）将会从数学处理方法中受益：它比那种含糊的、挥挥手式的解释要好。

有些科学家相信，简单地承认时间之箭就“对许多经济过程提供了新的看法”。另一些人认为，在非平衡态热力学中发现的演化规律，有助于领会人类社会进行与发展中的大尺度图像。拉·浩亚研究所的威斯特（Bruce West）和加利福尼亚索尔克疫苗研究所的索尔克（Jonas Salk）相信，把简并处理和综合处理两种方法融合在一起，“我们将会认识到决定人类行为和演化的大自然法则。这样做也就会增加我们的能力，使我们更好地了解人的本性”。

非线性动力学和非平衡态热力学所取得的成果，已经震撼了20世

纪的科学，并且对于时间给出了一个精深奥妙的重新评价。非线性方程向我们表明，热力学能够解释线性的时间，也能够解释循环的时间。当时间之箭毫不含糊地指向热力学平衡时，在向这一目的地疾进的过程之中，可能出现重复性的表现，例如化学钟里的颜色变化，黏菌发送出的一波一波的化学信息，或是人的心脏的跳动。我们也已经发现，这些非线性方程中包含着秩序和混沌两者的配方。在化学钟里，同样一组数学表述会产生规则的和无规则的两种颜色变化；混沌不是别的，正是自组织的一种不守规矩的形式。

化学钟不仅仅只是理想的、五彩缤纷的实验。化学钟也启示我们自然系统是如何成长和发展的。在人生的线性轨迹上——从胚胎、出生到死亡和腐烂，我们的细胞中含有的化学钟，柔和地调节着一系列维持生命所必需的循环。描述这些化学钟的数学分支不仅包含时间之箭，而且包含了生命的配方。

我们已经到了一个转折点。与其从原子和分子的抽象世界出发，去试图推导雪人和瓷器店公牛的宏观世界中的主导规律，为什么不反其道而行之？为什么不去相信我们的感觉所告诉我们的以及诗人们千百年来所描述过的一切？我们的世界中的基本现实，不是要在它最小的组成部分中发现，而是要在原子和分子的巨大集合的变化能力中发现，正如热力学第二定律所表述的那样。

这是一步大胆的充满想像力的跨越，但是当我们审视相空间图像时，我们已看到这步跨越是有良好基础的。这已向我们表明，简单的“可逆”机制如何产生出混沌和不可逆的行为。当一只摆动的摆锤的相空间图明显地限定为轨道上的一些点子，不再给出时间流逝的踪迹的时候，只要我们考虑其他更复杂的系统，我们就可以得出完全不同的结果。对于仅仅只有三个台球的情况，动力学表现就变得如此不稳定，以至于不可能预言它们的行为。相空间图变成黏乎乎的一团可能性，发芽抽出许多卷须，蔓延扩散开来，探索着无数的可能性。时间的流逝相应于这种绒毛的增长。

混沌的出现以及因此而出现的时间之箭，在当我们与我们周围的世界——其中物体含有无数个原子——打交道时，是如此清清楚楚地摆在我们眼前。当一个气体分子仅仅经过几次碰撞之后，我们就无法在任何肯定的程度上预言它的行为。对于更大质量的物体，例如绕日运行的一颗行星的轨道，在某些方面我们可以预言它在长得多的时间内的行为，甚至可达几百万年。但是即使如此，最终也会成为不可预测的。我们将只能够预言宇宙的大体性质。例如温度和密度。

这些关于复杂性的新想法，看起来会使传统科学颠倒过来。热力学第二定律，它描述巨大数目的分子的行为，是基本实际的一个部分。而简并主义，它企图单独用世界的微观成分的行为来解释世界，是无效力的。与其认为时间之箭是一种幻觉，我们倒应该问一下，这些时间对称的“基本”定律是否是近似、或者是幻觉。它们看来毕竟只适用于极其简单的体系。

时间的箭头好似在某个特定的长度尺度上出现，这种可能性我们不用再管了。我们可以看到，时间之箭如何在事物复杂化的过程中揭示了自己。只有在绝对简单的理想程度上，它的作用仍然是个谜。

下一步将发生什么呢？看来我们在朝一个新的视野探索。因为，如普里高津所说：“不可逆性不是对动力学定律的某种附加近似，而是把动力学置于一个更为庞大的形式体系之中。”绝不应该忘记，我们所知道的微观量子世界是扑朔迷离的——以测量问题为典型代表——它需要涉及一个宏观世界的预先存在，才使讨论具有一点点意义。并且，我们也已提到过，一个包括量子力学和广义相对论两者的更一般的理论，很可能是内禀时间不对称的。我们远不是已经接近对大自然彻底了解的尽头，像某些理论物理学家想使我们相信的那样；我们也许正处在一个崭新的大厦的入口处，在那里时间起的作用是主要的，而不是边缘的。

当前，多数科学家拥护简并派的观点。但是在我们贯穿整个科学

的旅行之中，我们已经看到需要把宇宙作为一个整体来考虑，并因此采用一种更为综合的观点。如作家托夫勒（Alvin Toffler）所作的评论："当代西方文明最高度发展的技巧之一是分析：把问题分解成它们的最小组成单元。我们精于此道并为此而津津乐道。但是我们常常忘记把这些离枝碎片再拼合回去。"需要有一个新的理论，它能够使我们对时间之箭作深入理解。从贯穿于此书始终的讨论中我们可以清楚地看到，这新的理论应当能够做些什么。它应当驱逐掉从基础上破坏爱因斯坦相对论的奇点。它应当结束量子论中有关测量手段作用的争论。它应当最终地宣布时间旅行、新生儿自谋杀以及白洞（黑洞的时间反演）是非法的，宣布基于主观和幻想的论证是无效的。

我们所得到的无法避免的结论是，由于过分强调非常简单的或者理想化的模型，使得物理学家们所采用的传统方法，即使对于解释日常现象也是太狭隘了。我们必须认识到现实世界的内禀复杂性，并接受一个根本性的概念更新。

今天我们的努力已经远远超过了玻耳兹曼当初的想法，然而他的渊博见解仍然和我们相伴。我们所栖身的这个世界，将来有着无穷尽的可能性，而过去却不可重得地留在我们的后面。时间之箭对于保持科学的完整性是必不可少的。它是创造力的手段，用它，生命才能够被理解。只有通过对这些事实的认识，我们才能在理性上开始沟通人文与科学两方面的经验。

# 上帝与新物理学

# 总序

科学，特别是自然科学，最重要的目标之一，就是追寻科学本身的原动力，或曰追寻其第一推动。同时，科学的这种追求精神本身，又成为社会发展和人类进步的一种最基本的推动。

科学总是寻求发现和了解客观世界的新现象，研究和掌握新规律，总是在不懈地追求真理。科学是认真的、严谨的、实事求是的，同时，科学又是创造的。科学的最基本态度之一就是疑问，科学的最基本精神之一就是批判。

的确，科学活动，特别是自然科学活动，比较起其他的人类活动来，其最基本特征就是不断进步。哪怕在其他方面倒退的时候，科学却总是进步着，即使是缓慢而艰难的进步，这表明，自然科学活动中包含着人类的最进步因素。

正是在这个意义上，科学堪称为人类进步的“第一推动”。

科学教育，特别是自然科学的教育，是提高人们素质的重要因素，是现代教育的一个核心。科学教育不仅使人获得生活和工作所需的知识和技能，更重要的是使人获得科学思想、科学精神、科学态度以及科学方法的熏陶和培养，使人获得非生物本能的智慧，获得非与生俱来的灵魂。可以这样说，没有科学的“教育”，只是培养信仰，而不是教育。没有受过科学教育的人，只能称为受过训练，而非受过教育。

正是在这个意义上，科学堪称为使人进化为现代人的“第一推

动”。

近百年来，无数仁人智士意识到，强国富民再造中国离不开科学技术，他们为摆脱愚昧与无知做了艰苦卓绝的奋斗。中国的科学先贤们代代相传，不遗余力地为中国的进步献身于科学启蒙运动，以图完成国人的强国梦。然而应该说，这个目标远未达到。今日的中国需要新的科学启蒙，需要现代科学教育。只有全社会的人具备较高的科学素质，以科学的精神和思想、科学的态度和方法作为探讨和解决各类问题的共同基础和出发点，社会才能更好地向前发展和进步。因此，中国的进步离不开科学，是毋庸置疑的。

正是在这个意义上，似乎可以说，科学已被公认是中国进步所必不可少的推动。

然而，这并不意味着，科学的精神也同样地被公认和接受。虽然，科学已渗透到社会的各个领域和层面，科学的价值和地位也更高了，但是，毋庸讳言，在一定的范围内，或某些特定时候，人们只是承认“科学是有用的”，只停留在对科学所带来的后果的接受和承认，而不是对科学的原动力，科学的精神的接受和承认。此种现象的存在也是不能忽视的。

科学的精神之一，是它自身就是自身的“第一推动”。也就是说，科学活动在原则上是不隶属于服务于神学的，不隶属于服务于儒学的，科学活动在原则上也不隶属于服务于任何哲学的。科学是超越宗教差别的，超越民族差别的，超越党派差别的，超越文化的地域差别的，科学是普适的、独立的，它自身就是自身的主宰。

湖南科学技术出版社精选了一批关于科学思想和科学精神的世界名著，请有关学者译成中文出版，其目的就是为了传播科学的精神，科学的思想，特别是自然科学的精神和思想，从而起到倡导科学精神，推动科技发展，对全民进行新的科学启蒙和科学教育的作用，为

中国的进步做一点推动。丛书定名为《第一推动》，当然并非说其中每一册都是第一推动，但是可以肯定，蕴含在每一册中的科学的内容、观点、思想和精神，都会使你或多或少地更接近第一推动，或多或少地发现，自身如何成为自身的主宰。

《第一推动》丛书编委会

# 致谢

我要特别向苏塞克斯大学的约翰·巴罗博士致谢，他对本书提出了详细的评论，使之有了很大的改进。本书所讨论的问题也成了我所在的系的同事们喝咖啡时热切讨论的话题，我发现与下列同事的谈话非常富有成果。他们是：斯蒂芬·贝丁博士，凯里·欣顿先生，J.普弗什博士，斯蒂芬·乌恩汶博士以及威廉·沃克先生。

本书作者及出版者希望感谢：Faberand FaberLtd惠许引用《盆栽天竺葵》中诺尔曼·尼克尔逊的《膨胀的宇宙》；Harvester Press Ltd惠许引用D-R.霍夫斯塔特的《哥德尔、埃舍尔、巴赫》及D.R.霍夫斯塔特和D.C.顿耐特的《精神的我》；Menthuen London Ltd惠许引用托马斯·吉尔比编辑的圣托马斯·阿奎那的《神学大全》，卷一：《基督教神学》；理查德.P.费恩曼惠许引用他的《物理定律的本性》；

Pergamon Press Ltd惠许引用由罗纳德·邓肯及米兰达·维斯顿-史密斯编辑的《活的真理》一书中赫尔曼·邦迪爵士的《宗教是件好事》。

# 前言

50多年以前，物理学中发生了奇怪的事。在从事这门科学的人当中，突然出现了许多有关空间、时间以及有关心与物的全新看法，这些看法稀奇古怪、令人惊讶。只是到了今天，这些看法才开始波及大众。那些使物理学家们好奇并激发了他们灵感的看法现在终于引起了公众的注意。公众在此之前从未想到，人类思想中已经发生了一场大革命。新物理学已经成年了。

20世纪头25年，有两个重大的理论被提出来，那就是相对论和量子论。20世纪的物理学大部分就是从这两论中产生出来的。但新物理学很快就显示出，它并不仅仅是物理世界的一个更好的模型而已。物理学家开始意识到，他们的发现要求对实在的最基本方面作出全新的界定。他们于是学着以完全出人意料的、全新的方法来解决问题，而这些方法似乎是违背常识的，它们更接近神秘主义而不是接近唯物主义。

这场革命的成果只是到了现在才有哲学家和神学家来采摘。许多普通的人，在寻找他们生命背后更深一层意义的过程中，也发现他们对世界的信念与新物理学很合拍。物理学家的看法甚至在心理学家、社会学家当中也获得了同情，特别同情物理学家的是那些主张对问题要作整体论研究的人。

我主讲了若干次关于现代物理学的讲座、讲演，发现人们越来越觉得基础物理学正指引人们重新评价人以及人在宇宙中的地位。宇宙是如何起始的又会如何结束？什么是物质？什么是生命？什么是精神？这些关于存在的深刻问题并不新鲜。新鲜的是，我们现在可能已

处在解答这些问题的边缘了。之所以会有这样令人吃惊的前景，是因为近来物理学出现了某些令人轰动的进展。此处所说的物理学，不仅是新物理学，而且也有其近亲——新宇宙学。

我们头一次可以对整个宇宙进行统一的描述了。宇宙是怎样诞生的？这个谜是科学的最基本而又最令人生畏的问题。没有超自然的输入，宇宙能诞生吗？有一古老的看法认为，“无中不可能生有”。量子物理学似乎已在这一论断上打开了一个缺口。物理学家们现在谈论着“自创造的宇宙”，即一个自发地爆发形成的宇宙，像有时在某些高能过程中从不知何处爆出的亚核粒子一样。这种理论的细节是否正确？这个问题并不怎么重要，重要的是，现在有可能为万有的创生设想一个科学的解释了。难道现代物理学已把上帝完全取消了吗？

本书不是谈宗教的书。本书谈的是新物理学对以前属于宗教的问题产生的影响。我尤其不想讨论宗教体验或道德问题。但本书也不是一本纯科学的书。本书是关于科学及其广泛含义的书。因而，本书不可避免地必须在这里或那里详细地解释一些技术性问题。但我并不想声称书中所涉及科学的讨论是成系统的，或完整的。读者不必害怕，书中不会有令人心烦的数学或一串串的专门术语来折磨你。我已尽可能地避免使用专门术语。

本书主要是面向普通读者的，包括无神论读者和宗教教徒读者，这些读者都是以前没有科学知识的。然而，我希望本书也包含有一些具有真正学术价值的材料。我尤其认为，最近的关于宇宙学的工作并没有受到哲学家和神学家的注意。

本书的主题是我所谓的四大存在问题：

为什么大自然的规律是现在这样的？

为什么宇宙是由现在组成它的各种东西所组成的？

这些东西是如何起始的？

宇宙如何获得了组织？

在本书将近结尾时，对这些问题的试探性答案开始显现。这些答案是以物理学家对自然的看法为基础的。这些答案或许全错了，但我相信，物理学的确是处在提供这类答案的独特地位上。我认为，与宗教相比，科学能为人指出一条更为确切的通向上帝的道路。或许这样的想法有些古怪。无论正确与否，科学实际上已进展到这一地步，它可以认真地解决从前是属于宗教的问题了。这一事实本身就显示了新物理学的深远影响。

尽管我自始至终地努力摒除我自己的宗教观点，但我对物理学的阐发不可避免地是属于我个人的。毫无疑问，我的许多同事会强烈反对我所试图引出的结论。我尊重他们的观点。本书不过是某一个人对宇宙的看法而已，对宇宙的看法除此之外还有很多。我之所以要写这本书，是因为我相信人在世上看到的东西很有限。

# 第一章　科学与宗教在变化着的世界中

有智慧的人将宗教和科学理论并用，以调节自己的操行。

J-B-S-哈尔旦因

因本宗教法庭命令我要完全放弃那种认为太阳是不动的中心的错误论点，并且禁止我以任何方式持有、捍卫或宣讲该错误信条，我发誓弃绝、诅咒并憎恶上述谬误及邪说，以及一切其他与上述教会相悖的谬误及教派。

伽利略

科学与宗教代表了人类思想的两大体系。对这个行星上的大多数人来说，宗教对他们处理事务是具有最显著的影响的。而科学与他们的生活的接触，却不是在心智层面上，科学实际上是通过技术影响到他们的生活。

尽管宗教思想在普通公众的日常生活中影响巨大，但我们的大多数社会事业机构是按照务实的方式组织起来的，其中即使有宗教，也只是把宗教淡化为一个程式化的角色而已。例如，英国教会在宪法中的地位就是这样的。也有些例外，如爱尔兰和以色列在法律意义上就仍然是宗教国家，而好战的伊斯兰教的复兴则正在增加宗教在政治和社会决策方面的影响力。

在工业化世界中，科学的影响及成就是最为明显的，因而，人们与主要的传统宗教机构的联系急剧减少。在英国，按时参加教堂礼拜的人占总人口的百分比很低。然而，假如就此得出结论，以为参加教

会活动的人数的减少可以直接归因于科学技术地位的提高，那可就错了。很多人在其个人生活中仍对可被划为宗教性的世界抱有很深的信仰，尽管他们或许已经放弃了或至少是无视了传统的基督教信条。任何一位科学家都会证实，即便说宗教已被逐出人们的意识领域，也不能说宗教空出的地盘肯定已被理性的科学思想所占领。因为科学正像任何一种排他性的宗教一样，尽管在实际的层面上对日常生活影响很大，但对普通公众来说，也是同样地令人无从捉摸，同样地难以领悟。

与宗教的衰落较为相关的原因是，科学通过技术使我们的生活发生了急剧的变化，以至传统的宗教或许显得失去了那种直接性，而那种直接性又是在人们应付现今的个人及社会问题时向他们提供现实的帮助所不可缺少的。假如教会在今天受到了很大程度的忽视，那也并不是因为科学在与宗教进行的长期战斗中取得了最后胜利，而是因为科学如此彻底地使我们的社会重新定向，以至圣经所描绘的世界图景现在似乎在很大程度上是缺乏时代性了。正如最近一位在电视上挖苦人的人所说，我们的邻居很少有牛或驴供我们贪恋了。[1]

世界上的各个主要宗教都是建立在公认的智慧和信条上的，它们都根植于过去，难以应付变化着的时代。教徒们匆忙之中发掘的灵活性，已使基督教能够容纳一些现代思想的新观念，以至今天的教会领袖在一个维多利亚时代的人眼里，很可能被认为是个异端分子。但是，任何一种以古代的概念为基础的综合性哲学都面临着适应太空时代的艰难任务。因而，很多幻灭的信徒转向了“次要的”宗教，这些宗教似乎与星球大战及微晶体的时代更合拍。与不明飞行物体、超感官知觉、心灵交往、信念治疗方法、超验冥想相关的崇拜以及其他基于技术的信仰大行其道就证明，在一表面上是理性的、讲科学的社会中，信仰和信条具有持久的吸引力。因为这些古怪的信念是属于毫不羞耻的非理性的，虽然它们都有一个科学的外表。用克里斯多弗·埃文斯1974年在Panther出版社出版的一本书的标题来说，这些信念都是

对“非理性的崇拜”。人们转向它们，并不是为了求得心智上的启蒙，而是为了在一个艰难而无常的世界中获得精神上的安慰。

科学侵入了我们的生活、我们的语言和宗教，但这并不是在理智层面上的入侵。绝大多数的人对科学的原则不理解，而且也不感兴趣。科学仍然是一种巫术，投在科学的实践者们身上的目光仍是恐惧与怀疑参半的。随便到哪家书店去看看，就可以看到关于科学的书通常都是归在“秘学”类下的，而现代天文学教科书则因书架不够，与《百慕大三角》和《众神之车》挤在一起。为了在我们现今的社会中建立秩序，人们对科学及理性思维的重要性倍加称道，但在个人层面上，大多数人仍然觉得宗教教旨要比科学论点来得更有说服力。

我们生活于其中的这个世界，尽管外表上很科学化，但骨子里仍是属宗教的。在像伊朗、沙特阿拉伯这样的国家里，伊斯兰教仍是占有优势的社会力量。在工业化的西方，尽管宗教已经分裂，多样化了，有时还变成了模模糊糊的伪科学性的迷信，但人们仍旧在继续寻求生命的更深的意义。这种寻求也不应该受到嘲弄，因为科学家自己也在寻求一种意义，他们想在宇宙的组成、运行的方式，生命的本质以及人的意识等方面获得更多的发现，以便能得出宗教信仰赖以形成的原材料。假如科学的测定揭示出地球的年龄为45亿年，那么，再去争论上帝创造天地是在公元前4004年或公元前10000年就没有什么意思了。任何宗教，假如其信仰的基础是可被证明为错误的假设，那么这种宗教就别想长命。

在本书中，我们将要看看某些基础科学的最新发现，并要探寻一下这些发现对宗教的含义。在很多情况下，我们可以看到，原先的宗教观念与其说是被现代科学所证伪，不如说是被现代科学所超越。科学家通过一个不同的角度观察世界，可以得出新鲜的见解，并且可以为人类及人类在宇宙中的地位描述出新的画面。

科学和宗教都有两个方面，即理智方面和社会方面。科学与宗教

的社会效应都还很不能令人满意。科学或许确实减轻了人类因疾病而遭受的苦难，为人类节省了诸多的劳动，提供了一系列发明，给我们带来了娱乐与方便。但科学也引出了用于大规模毁灭的可怕的武器，因而严重降低了生活质量。科学对工业社会的影响是好坏参半的。

在另一方面，有组织的宗教境况更糟。当然，没有人否认，在世界各地的宗教社区工作者中，有很多个别的无私奉献的例子，但宗教长久以来就已经制度化了。宗教常常更关注权力、政治，而不关注善与恶。宗教激情则更为经常地导致暴力冲突，搞乱了人类正常的宽容精神，释放了野蛮与残暴。基督徒在中世纪对南美当地人进行的种族灭绝就是较为可怕的例子之一，而在欧洲的历史上则到处都横陈着人类的尸体，那是些因为微小的教义分歧而被杀戮的人。即便是在所谓的启蒙时代，宗教仇恨和宗教冲突也遍布世界。具有讽刺意味的是，尽管大多数宗教都赞美爱心、和睦、谦卑，并将这些称之为美德，但世界上各大宗教组织的历史却常常是以仇恨、战争、傲慢为其特色的。

许多科学家对有组织的宗教持批评态度，其原因倒不是因为他们自己在精神上与这些宗教有什么不合，而是因为这些宗教所发生的影响在他们看来搞乱了正派的人类行为，在这些宗教卷入权力政治之争时尤其如此。物理学家赫尔曼·邦迪是个对宗教持严厉批评态度的人。他把宗教看作是“一种恶，一种严重的能够形成习惯的恶”。他举出当年欧洲对巫师进行疯狂大搜捕时的暴行为例：

在信仰基督教的广大欧洲地域，敬畏上帝的人常把被怀疑为巫婆的妇女烧死。这些人觉得这是他们应该勉力完成的职责，这职责是圣经明明白白地分派给他们的。事情很清楚，首先，是信仰使得那些本来正派的人干出了难以名状的可怕的事，这表明了人类日常的善心和对残暴的嫌恶何以而且也确实为宗教信仰所压倒。再者，这揭示了所谓宗教确立了绝对的、不变的道德根基的说法是完全不实的。①

邦迪声称，教会以及其他的宗教机构多少世纪以来所掌握的残暴的权力，已使它们在道德上名誉扫地了。

尽管宗教在四处自夸，但没有人会否认宗教仍是社会中最能引起分裂的力量之一。不管信仰虔诚的人们主观愿望是多么善良，血迹斑斑的宗教冲突史说明，没有什么证据能够证明在各大有组织的宗教里含有人类道德的普遍准则。而且，我们也没有理由相信，那些不属于这些宗教组织的人就是缺乏爱心和热心，或就是彻头彻尾的无神论者。

当然，并不是所有的信教的人都是宗教狂热分子。绝大多数基督徒今天都嫌恶宗教冲突，而且也为教会在过去卷入酷刑、凶杀、镇压而感到痛惜。但是打着上帝旗号的野蛮暴力时常爆发，触目惊心，至今仍袭扰着社会。这并不是宗教的反社会方面的唯一表现。在所谓的文明开化的国家如北爱尔兰和塞浦路斯，人们仍旧因为宗教原因而在教育甚至是居住方面实行隔离。即便是在自己的内部，宗教组织也常常准许歧视的存在，歧视妇女、少数民族，歧视同性恋者或宗教组织的领袖所认定的任何劣等人。我觉得，天主教及伊斯兰教中妇女的地位，以及南非教会中黑人的地位特别令人难以忍受。尽管许多人在他们的宗教被形容为罪恶、褊狭时会被吓得要死，但他们却很痛快地一致认为，世界上的其他宗教都是作恶多端的。

宗教组织一旦制度化了，法制化了，宗教偏执的伤心惨目的历史似乎就注定不可避免了。而且宗教偏执行为在西方世界引起了人们对国教的极大的不满。许多人正转而投奔所谓的“次要”宗教，以图找到一条获得精神满足的途径，既不那么刺耳。却又那么温柔。当然，现在各式各样的新宗教运动多的是，其中的某些运动比传统的宗教更褊狭，更邪恶。但是，现在有很多人强调神秘主义以及安宁的内心探索的重要性，并以此与传道的狂热相对，因而吸引了对定为国教的宗教的社会及政治影响持批评态度的那些人。

关于宗教的社会方面就谈这些。那么，宗教的理智内容又是什么呢？

在人类历史的大部分时期，男男女女们之所以皈依宗教。并不只是为了寻求道德指引，而且也是为了寻求关于存在的基本问题的答案。宇宙是如何被创造的？宇宙又会怎样终结？生命和人类的起源是什么？只是到了近几个世纪，科学才开始为这类问题的解决做出自己的贡献。科学随之与宗教发生了冲突，这些冲突都被详细记录下来。由伽利略、哥白尼、牛顿打头，随后来了达尔文、爱因斯坦，直到计算机和高技术的时代，现代科学对很多根深蒂固的宗教信念进行了阐明，这些阐明是冰冷的，有时是具有威胁性的。于是，这就造成了一种感觉，觉得科学与宗教是天生的死对头。这种看法得到了历史的鼓励。早期的教会试图顶住科学的闸门，阻止科学进步的洪水滔滔而下，造成的结果是从事科学的人从此对宗教产生了深深的怀疑。科学家们毁坏了很多人所珍视的宗教信念，于是他们被很多人看成是信仰的破坏者。

然而，科学方法所获得的成功是毫无疑问的。物理学，这科学之王，为人类的理解开辟了几个世纪以前想都没有想到过的新路。从原子内部的作用到黑洞的玄妙，物理学使我们得以理解自然中某些最隐秘的秘密。而且也使我们得以控制我们的环境中的许多物理系统。科学推理的巨大力量使每一日都有许多现代技术的奇观得以证明。那么，对科学家所持有的世界观也抱有一些信心，似乎就是顺理成章的了。

科学家与神学家是从完全不同的出发点出发，对有关存在的深刻问题进行探索的。科学以仔细的观察、实验为基础，使理论得以建立，将不同的经验联系起来。科学寻求的是自然在运行中的规律性，这些规律性很有希望地揭示了制约着物质和力的基本规律。科学对存在的问题进行这样的探索，其中的关键是，科学家假如遇到了他所持有的理论的反证，他就得放弃那个理论。尽管个别的科学家或许会顽

固坚持某个他们所珍视的观点，但科学界作为一个群体是随时乐于采纳新观点的。科学原则之争不会引发人类相互残杀的战争。

与科学相比，宗教是建立在启示和公认的智慧的基础上的。声称包容了不可更改的真理的宗教信条是难以做出修正以适应变化着的观念的。真正的信徒必须坚持自己的信仰，不管有什么明显的反证。这“真理”据说是直接传达给信徒的，没有经过集体调查过程的筛选、提炼。天启“真理”的麻烦之处在于，这样的真理有可能是错的，即便它是对的，其他的人也需要想好了之后才能赞同信徒们的信仰。

很多科学家对天启的真理持嘲笑态度。其中有些科学家甚至认为所谓的天启真理是明白无误的一种恶：

一般说来，一个受到启示的信徒的思想处于可怕的自大状态。处在这样的状态之中，他就能说这样的话：“我知道，我明白，而那些跟我信仰不同的人是错的。”这样的自大心态在宗教领域最为普遍，而且只有在宗教领域里，人们才对他们的“知识”觉得有如此这般彻底的把握。在我看来，一个人竟能这么自我感觉良好，竟自觉这么出类拔萃，而不把所有那些与他们信仰不同的人放在眼里，这样的一个人是很令人讨厌的。这本来就够糟的了，但很多信徒还竭尽全力传播他们的信仰，至少是向他们自己的孩子传播，但也更经常地是向别人传播（而且，在历史上，有很多靠武力和惨无人道的野蛮传播信仰的好例子）。明摆着的事实是，最诚实的人以及所有智力水平高低不同的人都各不相同；而且宗教信仰一直是不同的。因为最多只有一种信仰是真实的，那么，这自然就是说，人类极其可能在天启宗教方面坚决而诚实地信仰了某种不真实的东西。人们本可以期望这一明显的事实能够引出某种谦卑，使信徒多少想到不管一个人的信仰有多么深，一个人很可能是信错了的，但任何信徒都没有这基本的谦卑。信徒不把自己的信仰强行灌输到他所掌管的人的脑子里便不肯罢手（这种情况在现今的发达国家一般限于信徒对其子女）。②

然而，那些有过宗教体验的人总是把他们自己获得的启示看作是比多少科学实验都坚实的信仰基础。不错，很多职业科学家也笃信宗教，而且显然并不觉得让他们的人生观的这两个方面和平共处有什么智力上的困难。问题是如何将许多根本不相同的宗教体验变成一连贯的宗教世界观。例如，基督教的宇宙论就与东方的宇宙论差异巨大。至少，两者之中有一个是错的。

但是，若从科学家对天启真理的怀疑推开去，推出科学家是冷酷的、不通人情的、工于计算、没有灵魂的人，只对事实和数字感兴趣，那可就大错特错了。新物理学兴起的同时，科学家对科学的更深的哲学含义的兴趣也有了巨大增长。这种兴趣是科学工作不那么广为人知的一个方面，而且常常是完全突然地产生的。在筹划一部关于精神和超自然现象的电视系列片时，病理学家、作家、电视制片人基特·派德勒这样描述了他见到现代物理学家们对超出物理学的问题如此关心时所感到的惊讶：

几乎整整20年来，我作为一个快乐的信仰还原论的生物学家进行我的研究，以为只要我努力研究，最终就能揭示出基本的事实。后来。我开始阅读新物理学的书。结果，我以前的信仰被粉碎了。

作为生物学家，我以前认为物理学家都是头脑冷静、清晰、不易动感情的男男女女，他们以一种冷静的、局外人的眼光居高临下地看待自然，把落日的光辉分解成波长和频率；他们是一帮观测者，把结构精巧的宇宙撕成死板的形式成分。

我的错误是巨大的。于是，我开始研读传奇般大人物的著作——爱因斯坦，玻尔，薛定谔，狄拉克。我发现，他们并不是冷静的局外人，而是一些富有诗意、笃信宗教的人。他们所想象的东西是那么巨大，那么新奇，以至我所谓的“超自然的东西”相形之下显得几乎平淡无奇了。③

具有讽刺意味的是，一直走在各种学科之前的物理学现在正对精神越来越倾向于肯定；而生命科学则仍旧走在19世纪的物理学的路上，现在正试图完全取消精神。心理学家哈罗德·莫洛维茨对物理学和生命科学如此转换对精神的看法提出了如下的评论：

实际情况是，生物学家们从前认为人的精神在自然界的分类等级之中占有一个特殊的地位，现在则义无反顾地走向赤裸裸的唯物论，而19世纪的物理学就是以唯物论为其特色的。与此同时，面对着咄咄逼人的实验证据的物理学家们则脱离严格机械的种种宇宙模型，转而把精神看作是在一切物理事件中扮演着一个与事件不可分离的角色。这两个学科就像是坐在两列逆向飞驰的火车上的乘客，彼此都没注意对开过来的火车上正发生什么事。④

在下面的几章里，我们将要看到新物理学如何赋予“观察者”在物理实在的自然中的中心地位。越来越多的人认为，基础科学近来的进展更有可能揭示出存在的更深一层的意义，而不是投合传统的宗教。总之，宗教对这些科学进展不可能视而不见。

# 第二章　创世

起初上帝创造天地。

《创世记》第1章第1节

但当时没有人在现场看。

斯蒂芬·温伯格《最初的三分钟》

有天地创生这回事吗？假如有，那又是什么时候发生的？原因是什么？存在之谜最深奥，最难猜。大多数宗教都有涉及万物如何起始的说法，现代科学也有。在本书中，我将借助宇宙学的新近发现来猜测这创生之谜。本章讨论的就是宇宙总体的起源。有人用“宇宙”（universe）这个词来表示太阳系或银河系。但我将按“一切物理性的存在的东西”这一较为常规的意义使用“宇宙”一词，我所说的宇宙是散布在一切星系之中、之间的一切物质，一切形式的能量，一切非物质的东西如黑洞、引力波以及一切延伸向无限（假如果真如此的话）的空间。有时，我将用“世界”（world）来指上述的一切。

任何声称对物质世界提供某种理解的思想体系都必定要对世界的起源说一些话。在最基本点上，有两种泾渭分明的说法。宇宙要么是一直存在着（以这样或那样的形式）；要么就有起始，多少有些突然地起始于过去某一特定的时刻。这两种说法长久以来一直困扰着神学家、哲学家、科学家，而且这两种说法对普通人来说也明显地难以理解。

假如说宇宙在时间上没有起源，就是说宇宙一直存在，那么，宇

宙的年龄就是无穷大的了。无穷大这个概念让许多人眩晕。假如早已有了无穷多的事件，为什么我们发现我们自己是生活在现在这一时刻？难道宇宙在永恒的时间里一直是静止的，只是到了相对晚近的时候才突然活动起来？或者，在永恒的时间里，一直存在着某种活动？另一方面，假如说宇宙是有起始的，那么就是承认了宇宙是突然从虚无中生出来的。这似乎暗示着有一个最初的事件。假如真是有，那最初事件的起因又是什么？这样的问题还有意义吗？

许多思想家面对这些问题畏缩不前，转而去寻求科学的证据。科学在宇宙起源的问题上能告诉我们些什么？

如今，大多数宇宙学家和天文学家都支持约在180亿年前确有天地创生的理论。当时，物质的宇宙在一场可怕的爆炸中产生了。那场爆炸就是现在人人皆知的“大爆炸”。“大爆炸”理论是很惊人的，但这一理论得到许多证据的支持。人们接受该理论的所有细节也好，不接受也好，其中的关键假说，即过去有某种创生的假说，从科学的观点来看是有力的。该理论之所以看上去令人信服，其直接理由来自热力学第二定律这一已知的物理学中最为普遍的定律所包含的庞大科学证据。就其最广泛的意义而言，热力学第二定律认为宇宙的无序程度与日俱增。宇宙在缓慢地、但却是无可挽回地坠人混沌。热力学第二定律的例子到处都是——建筑物倒塌，人变老，山脉跟海岸线受侵蚀，自然资源被消耗。

假如一切自然的活动都导致了更多的无序（以某种适当的方式测量出来的），那么，世界就一定是在不可逆转地变化，因为要使宇宙回到昨日的状态，就意味着把无序降低到先前的水平上，而这一点又是跟热力学第二定律不相容的。然而，乍一看，世上似乎很有些第二定律的反例。新建筑起来了，新结构生长了。难道每一个新生的婴孩不都是无序中生出有序的例子吗？

在这些例子中，你得注意你所观察的是整个系统，而不仅仅是你

所关心的事。在宇宙某一区域的有序增加，总是以另一处无序的增加为代价的。就以建筑一幢新大楼为例吧。建筑所用的材料不可避免地消耗了世界的资源，同时，在建筑的过程中耗用的能源也是不可回收的。做一下结算，就会发现总是无序为多。

物理学家们已发明了一个叫作熵的数学量来给无序定量。很多精心的实验也证明，一个系统中的总体的熵从来不会减少。假如一系统与其环境隔离开了，那么，该系统内的任何变化都会无情地使熵值增大，最后达到大到不能再大的程度。从此以后，就不会再有任何变化，整个系统就会达到一种热力平衡的状态。一个装有混合化学物质的匣子便是个好例。各种化学物质发生反应，或许由此产生一些热，各种物质改变它们的分子形式。诸如此类的一切变化都增加了匣内的熵。最后，匣中的物质取得了它们最终的化学形式，彼此的温度也一致了，于是，也就不会再有什么变化了。要想把最后的生成物还原到先前的状态也不是不可能，但这就得把匣子打开，再耗用能量和物质以便让已发生过的变化再变回去。这个操作过程的确会降低匣内的熵，但同时会产生更多的熵。

假如说宇宙真有一个有限的有序量，而且是在不可逆转地向着无序变化，最后达到热力学平衡，那么，就可以立即得出两个很重要的推论。推论一是，宇宙最终将会死去，消沉在它自己的熵里。这个推理，物理学家们称之为宇宙的“热死”。推论二是，宇宙不可能已有无限的过去，假如真有，那么宇宙就会在无限远的时间以前早就达到了最终的平衡状态。结论：宇宙并非过去一直存在。

我们在周围一切习见的系统中就可以看到热力学第二定律的作用。例如地球就不可能有无限的过去；假若有的话，那它的内核早该凉下来了。根据放射研究，地球的年龄可确定为45亿年。这跟月球以及各种陨石的年龄差不多。

而太阳则肯定不会热热闹闹地一直燃烧下去的。它的燃料储备年

年在减少，故而最后它会变凉、变暗。反过来推理，则太阳肯定只是在一段有限的时间以前才燃着的，因为它没有无限的能源。根据估算，太阳的年龄只是比地球的稍大一点儿。这估算与目前认为太阳系当初是作为一个单独的单位形成的这一类理论很符合。然而，太阳系只是宇宙的一个小山的组成部分。只是根据地球与太阳的情况就得出概括宇宙的结论也未免鲁莽。但太阳的确是一颗典型的恒星，光是我们所在的星系里就有成百亿其他的恒星，它们的生命史可供天文学家们研究。现存的恒星都各自处于其演化史上的不同阶段，这就使我们能够勾勒出恒星的诞生、演化、死亡的详细图画。

恒星与行星一道形成，形成的原因是，星际间主要由氢构成的大块稀薄的气云在缓慢地收缩，分裂。今天，要想找到星系中哪些区域有恒星正在诞生是很容易的。这样的区域之一是猎户星云，肉眼就看得见。恒星并非是一下子形成的。例如，我们的太阳现在约50亿岁，然而，它的年龄充其量至多只是银河系中最老的恒星的一半。而太阳系的形成可以看作是银河系中发生了几百亿次的一个持续不断的过程所得出的一个结果而已，而这个过程将持续下去。于是，我们似乎可以说，就恒星和行星的形成而言，是谈不上什么创生的，不过是某种宇宙装配线不断地把原料（氢、氦以及很少的重元素）装配成恒星、行星罢了。

就算是一批恒星在不断地燃尽，另一批恒星又不断地形成并取而代之，难道这生生灭灭的循环从无限远的过去就一直是这么周而复始直至今日的吗？不。热力学第二定律告诉我们，不会有这种周而复始以至无穷的循环。恒星燃尽的物质永远也不会完全进入再循环。燃烧所用的能量，以星光的形式消散在太空里，要在太空中发散几百亿年。还有一些恒星的物质进入了黑洞，再也出不来了。

那种认为整个宇宙系统并非一直就是这么循环往复的观点，还有一个更为直接的理由。艾萨克·牛顿这位现代科学的奠基人之一证实，引力是一种普遍的力，作用于宇宙中一切物体之间。每个恒星，每个

星系，都因引力而互相靠近。因为天体是在太空中自由漂浮的，看来它们没有任何理由不因这无所不在的引力而会聚在一起。在太阳系里，众行星因引力塌向太阳的事之所以没有发生是因为有离心力——各个行星都在绕太阳旋转。同样，银河系也在旋转。但现在我们还没有证据证明整个宇宙也在旋转。各个星系显然不能永远挂在太空中，所以，宇宙也不可能永远是目前这个样子。

尽管这个宇宙之谜自牛顿以来就被意识到了，但只是到了20世纪20年代才发现了它的解答。美国天文学家埃德温·哈勃发现，各个星系不是在相聚，而是在以很高的速度相离。哈勃注意到，各星系的光在颜色上稍稍有些异常（用术语来说就是“红移”），这说明各个星系在急速倒退。因为光是由不同的光波组成的，因此，一个移动的光源能将光波拉长或缩短，正如一辆行进中的车辆会把它发出的声波拉大或缩短一样。一辆汽车的发动机的声音或一列火车的汽笛声，在该汽车或火车急速开过我们身边时，其音高会降低很多。这里，只要把“音高”代换成“颜色”，就得到了哈勃红移。只是这里所说的各行星的退行速度比汽车或火车速度大得多。遥远的星系每秒退行成千上万英里。

哈勃的发现有时被人给以错误的解释，被认为是说明了我们所在的星系是众星系的中心，其他所有的星系都在飞离我们。这种看法是错误的。因为离我们远的星系退行速度要比近处的星系退行速度大，因而远处和近处的星系彼此间的距离也在扩大，所以，事实上每一个星系都在脱离其他的星系。这就是有名的“宇宙膨胀”。无论你从宇宙中的哪个位置上观察，星系扩散的图式看上去都是很一致的。

宇宙膨胀的概念与现代人关于空间、时间、运动本质的思想十分吻合。阿尔伯特·爱因斯坦在科学界中地位之高，有如圣保罗在基督徒中的地位，他那让人绞尽脑汁方能理解的相对论，极大地改变了我们对时空及运动的看法。尽管爱因斯坦的空间弯曲及时间弯曲的概念，耗时60年才进入普通大众的想象之中，物理学家们却早就接过他的时空弯曲的概念来解释引力了。

引力为一切大规模的宇宙现象提供了动力。在天文尺寸的物体中，引力要比其他的力如磁力、静电力大得多。形成星系并控制星系间的运动的，是引力。要解释宇宙的膨胀，引力就是关键。

爱因斯坦令人信服地说明，引力使空间、时间伸长或扭曲，而且这可以通过观察太阳的引力使擦过其表面的星光弯曲这一现象得到直接的验证。从地球上看，太阳背后的天空有很小的但是是明显的弯曲。时间的弯曲也可以得到证明，最直接的证明就是让钟表在空间飞行。时间在没有引力的环境中要比在地球的表面走得快。

假如太阳能拉长空间，那么，星系也能。因为星系是由很多太阳一样的恒星组成的。因而，天文学家们认为星系不是在空间中离散，而是星系间的空间在伸长。假如星系间的空间在“膨胀”，那么，每个星系的活动余地都在逐日增大。这就是说，宇宙在扩展，却不必扩展到某种外部的真空之中。

既然很多人觉得时空伸缩的概念难懂，我们就先把这些难懂的概念搁置一会儿。很明显的事是，既然宇宙在越变越大，那它以前一定要比现在小。假如过去一直就是目前这样的膨胀速度，那么，200亿或300亿年前，整个可以观测到的宇宙就会蜷缩进一个令我们感到十分陌生的小球中，在其中我们认不出任何天体。事实上，天文学家们已经发现，宇宙膨胀的速度正在变小，这就是说，那种高度压缩的状态事实上是在更早些时候发生的，距今大约是150或200亿年前（可与太阳的年龄——50亿年——做一比较）。因为当初的膨胀速度比现在的大得多，所以，星系离散在早期的时候，就像是场爆炸，而不是慢吞吞的膨胀。

有时人们说，我们今天所认识的宇宙当初是由于一种原初的“卵”发生爆炸而形成的。各星系都是当初爆炸的碎片，这些碎片现在仍在空间飞扬。这幅图景颇能反映一些真实的情况。但有时也会使人误入歧途。那个爆炸的东西之所以曾处于收缩状态，是因为空间处

于收缩状态。那种认为“卵”是包裹在真空中的想法是错误的。卵有壳、有核。然而，天文学家们则认为，宇宙既没有边或壳，也没有任何具有特权的中心。

我们现在算是跟“无限”这个微妙的问题缠上了。对那些疏忽的人来说，这个问题充满了陷阱。这个问题不但对于讨论膨胀的宇宙来说很重要，而且对科学和宗教这一类更为广泛的问题来说也是很重要的，因此，偏离一下我们的主题，讲一讲这个问题还是值得的。

科学家们长久以来就认识到，需要将他们关于无限的一切思考都建立在严格推导的数学步骤上，因为测量无限会导致各种各样的佯谬。例如，我们可以想想伊利亚的芝诺（公元前5世纪）提出的有名的“兔子和乌龟”的佯谬。在一次赛跑比赛中，乌龟开始领先，但兔子跑得快，很快就追上了乌龟。显然，在赛跑过程中的每一时刻，兔子和乌龟都各自处在一个位置上。因为两者都跑一段相同的时间——即时刻的数目相同——所以，按理说，它们通过的地段是相同的。但是，兔子要想追上乌龟，就必须在相同的时间里跑更远的距离，于是就要通过更多的地段。那么，兔子怎么能追上乌龟呢？

这个佯谬是归于芝诺名下的几个佯谬之一。要想求得这个佯谬的解答，就得仔细地陈述无限的概念。假如时间和空间是无限可分的，那么，兔子和乌龟就要跑无限段的时间、空间，在这里，“无限”的关键是，无限的一部分与无限的整体是一样的。尽管乌龟跑过的路要比兔子的短，可乌龟跑过的路段仍是与兔子的一样多（即无限），但我们知道，兔子不但跑过了乌龟跑过的路段，而且还跑了更多的路段！

研究无限的问题，就会弄出很多这一类令人吃惊的事来。数学家们花了许多个世纪进行逻辑论证，以便全面理解正确处理无限问题的规则。奇特的是，世上的无限不止一种。有东西的无限，可用整数代表（1，2，3，……，无穷大），还有一种更大的无限，用全部的整数都不足以表达出来。

说到几何学，人的直觉能把人搞得很迷糊。就以包围着一块给定面积的土地的篱笆为例吧。很容易看出，在面积已定的情况下，细长条形的土地需要的篱笆要多于正方形的土地。圆形土地需要的篱笆最少。但是，这块面积已定的土地周边到底能有多长？图1是一形状奇特的周边，是一次次地在三角形上再开三角形造成的。每开一次三角形，篱笆的周边都增长一些，它所圈的面积都要增多一点。但那周边永远也不会伸到外接圆之外去，所以，它所圈的面积永远是有限的。然而，随着在周边上开三角的次数的增多，周边就可以无限延长。这样，我们就能够想象，一无限长的篱笆围着一块面积有限的土地。



这一切与宇宙的创生又有什么关系呢？首先，这一切表明，“无限”之类的概念不应滥用，否则会导致无意义。第二，这一切表明，人们所获得的问题的答案常常跟直觉和常识相悖。这是科学给人的大教训之一。人们常常必须要依靠抽象的东西，依靠正规的数学操作来理解世界。平常的经验有时是不可靠的向导。

宇宙的体积是无限大吗？假如空间的体积是无限大，我们就可以想见有无穷多的星系散布其间，而且其密度也大体一致。于是，很多

人就不能理解，一个无限大的东西怎么还能膨胀？它要膨胀到哪里去呢？这个问题不难回答：无限的体积可以膨胀，但体积不变。（请记住“乌龟”教给我们的东西吧。）但是，当我们把这个模型拿来比附“宇宙卵”时，就出现了形象化方面的问题。假如到处都有星系，那么，从前就不可能有过一个体积有限，又有外壳，而且外壳之外没有物质的宇宙卵。因此，卵的比喻行不通。

设想一下，有这样一个无限的宇宙，这宇宙是一球体，包容了极大的体积的空间，其中有很多星系。现在再设想空间到处都在急速收缩，就像奇境中的爱丽丝吃过魔饼之后的情况一样。于是，该球体的半径就会越缩越小，然而不管怎样缩小，总是有无尽的空间，而且其外还有无限多的星系。假如该球体真是缩小到了无限小，那么，这就产生了一个从数学角度讲是棘手的问题，即一个无限缩小的无限宇宙。这个球体宇宙仍是没有中心也没有边界的，然而，任何这一类的球体不管当初是多么地大，经过无限缩小之后，其中所有的星系便会被压缩成一个点。天文学家们认为，宇宙发生爆炸之前，就处于这种无限收缩却无界的状态之中。

事实上，还有一个可用的宇宙模型可以避免诸多无限的纷扰。这个模型是爱因斯坦本人在1917年提出的。在确认空间能够弯曲这一事实之后，爱因斯坦说，空间能以许多种人所意想不到的方式把自己连接起来。可以拿地球的曲面为类比。地球的表面，面积是有限的，但它是无界的，一个旅行者走到哪里都不会碰上地球的边缘或边界。同样，空间也可能体积有限却没有边界。这种奇异的图景怕是没几个人能真正看明白，但数学自会替我们把其中的细节勾画出来的。该形体叫作超圆体（hypersphere）。假如宇宙是一超圆体，宇航员原则上就可以像麦哲伦一样作环宇航行了。他可以驾着火箭向同一个方向飞，一直飞到他的出发点。

尽管爱因斯坦所提出的超圆体宇宙是有限的，但却没有中心，也没有边界（正如地球的表面也没有中心，没有边界），因而，当它收

缩时，也不像一只宇宙卵。我们可以想象这超圆体缩至无限小，体积消失，就像一个圆面缩小到了半径为零一样（见图2）。

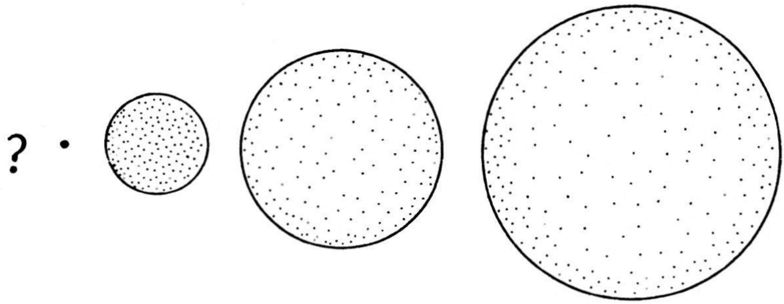



图 2　假如三维空间用一个二维的面来表示的话，那么，就可以有一个膨胀宇宙的模型，看上去很像是从无限小膨胀起来的气球。在这个模型里，空间是有限的，但也是无界的，在该空间中的一个观测者可以无障碍地漫游该空间。其中的小点代表星系（或星系团）。随着宇宙的膨胀，空间也拉长了，因而所有的小点都与其近邻的小点拉开了更大的距离。处在任何一个小点上的观测者都会看到，其他的所有小点都在有条不紊地后退，而且他还会觉得他所在的小点是其他小点作四散运动的中心。

对能伸能缩的空间进行的研究，使宇宙学家们提出了一个在细节上与圣经大不相同的宇宙创生的理论。他们所提出的科学理论最惊人的一点就是，他们认为空间本身，而不仅仅是物质，也是在大爆炸中创生的。假如你把图2正在膨胀的气球，即从无中膨胀起来的气球倒过来看，把它看成是正在收缩的气球，并以此作为宇宙的模型，那么，你就算是大致知道现代物理学是如何讲述宇宙创生的了。这里的一个重要论点是，我们这里所谈的宇宙不管是爱因斯坦的超圆体（即气球模型），还是体积无限的宇宙，要想设想空间无限收缩之后是什么样子是不可能的。大爆炸发生的最初那一刻，空间是无限收缩的，那一刻就代表着空间停止存在的时间边界。物理学家们把这边界称作奇点。

空间原来竟是从无中生出来的，这个离奇的想法是很多人觉得难以理解的，因为他们早已习惯把空间看成是“无”。然而，物理学家则把空间看成是一种能伸缩的媒介，而不是空空如也的“无”。的确，在下面的几章里我们就可以看到，由于量子效应，即使最纯的真空也能引起活动，其中充满了转瞬即逝的结构。对物理学家来说，“无”意味着“没有空间”以及没有物质。

离奇的事还多着呢。空间是与时间分不开的，于是，空间能伸能缩，时间也能伸能缩。大爆炸代表着空间的创生，同样，也代表着时间的创生。空间和时间都不能存在于奇点之前。简单地讲就是，时间本身也是从大爆炸之时起始的。

这一大堆稀奇古怪的观点只能够求助于数学才能完全把握。人的直觉是力不从心的向导。这也是科学方法之所以取得成功的主要原因之一。科学以数学为语言，就能够描述完全超乎人类想象力的东西。而现代物理学的大部分内容也的确是人的想象力所不及的。若没有数学所提供的抽象描述，物理学永远也不会超越简单的力学。当然，物理学像每个人一样，头脑里也有一些原子的模型，光波的模型，膨胀宇宙的、电子的以及其他诸如此类的东西的模型，但是，这些头脑中的模型常常是很不精确、十分离谱的。事实上，从逻辑上讲，任何一个人要想能正确地想象某些诸如原子之类的物理系统都是不可能的，因为那些物理系统包含着一些纯属我们经验世界之外的东西（第八章谈量子论时，我们就会明白这一点）。

人的想象抓不住实在的某些关键的东西，这就警告我们，我们不能期望通过日常经验获得一些简单的关于空间、时间和物质的概念，就以为某些重大的宗教真理（例如创世的本质）有了依据。

人们一直难于从理智上解释时间的起源。这并不是个新问题。亚里士多德、圣托马斯·阿奎那都认为，时间不是被创造的东西，否则就是说有第一事件了。第一事件的原因是什么？无，因为第一事件就是

说没有在它之前的事件。

时间的有限其实并不一定是意味着有一个第一事件。我们可以设想一些用数字代表的事件，其中零代表奇点。奇点并不是个事件，而是一种无限致密的状态或类似的东西，在奇点处时空不再存在。如果有人问："奇点之后的第一事件是什么？"这就相当于问"比零大的最小的数是什么？"世上并没有这样一个数字，因为所有的分数，不管多小，都总能被2除。同理，也没有第一事件。

若说时间无限，也同样让人为难，正如伊曼纽尔·康德后来所强调的那样：

假如我们设想世界在时间上是没有开始的，那么，对应于每一给定的时刻，都有无限的时间已经过去，而且世界上已经过去了相续的事态的一个无限序列。序列的无限就是它永远不能通过连续的综合而被完成。于是，这就是说，无限的世界序列已经过去是不可能的，因而世界有一个开始是世界存在的必要条件。①

我们必须记住芝诺给我们的教训，在处理无限的时候要小心谨慎。按康德的推理，兔子"通过连续的综合"永远也不能跑完无限序列的步数去追上乌龟。然而我们都知道，兔子肯定会追上乌龟。但是，要是说芝诺佯谬中的过去的时间是有限的，而康德所指的是无限的延续时间的过去，这也不能驳倒康德。芝诺和康德所涉及的都是无限的时刻。任何一个数学家都能证明，无限时间中的时刻并不比一分钟当中的时刻为多。芝诺和康德也都涉及了一个无限的数目，而这无限的数目并不能靠着"无限的延伸"而增大。

人们对康德的推理的另一个驳难是，他们认为时间是"流逝"的，这就是说时间是流动的、运动的。没有几个物理学家会承认时间真的是流动的或运动的。时间与空间一样，只是存在而已（我们在第九章要再谈这个话题）。

总而言之，一派人认为宇宙是无始无终的，另一派人认为宇宙的年龄是有限的，其过去的边界是奇点。这两派人的观点似乎都没什么大错。假设后一派的观点是正确的，这是不是就是说科学支持圣经关于天地创始的说法呢？

基督徒之间对圣经所描述的创世的真实性也没有一致的看法。1951年，教皇庇护十二世在罗马就现代科学宇宙学的问题[②]对主教科学院发表了讲话，拐弯抹角地提到了大爆炸理论，说“一切似乎都在表明，宇宙在有限的时间里有一个宏伟的开端。”他的话引发了激烈的反应（主要是神职人员的反应，科学家们的反应倒在其次）。当代的神学家们仍在争论不休，定不下宇宙创生之初的大爆炸是否就是《圣经》作者们据说受到启示而描述出来的上帝创造天地。美国圣母大学的厄南·麦克穆林最近写过一篇文章，题目是“宇宙学与神学应是何等关系？”其结论是，“首先，我们不能说基督教神创造天地的信条‘支持’大爆炸模型；次之，我们不能说大爆炸模型‘支持’神创造天地的信条。”[③]然而，当今很多百姓因受到要他们把大部分旧约看作是虚构的压力，现在看到现代科学的宇宙学给创世之说带来了显明的证据，于是感到快慰。

如果我们认为空间和时间就是在大爆炸中从无中爆发出来的，那么，就显然有“创世”，宇宙显然是年龄有限的。于是，热力学第二定律的佯谬也就立刻解开了。宇宙之所以还没有达到热力学平衡，是因为宇宙的无序化过程才进行了约180亿年，这过程离着完成还远着呢。而且，我们现在也可以明白，所有的星系为什么没有聚集起来。爆炸的力量把众星系炸散了，尽管现在它们离散的速度正在减小，但众星系重聚的时间还没有到来。

假如大爆炸理论只是以哈勃和爱因斯坦的工作为支撑，那么，该理论就不会得到它今天所得到的广泛支持。幸运的是，还有一些有说服力的证据。

宇宙诞生时惊天动地的力量爆发肯定会在宇宙的结构上留下某些印迹，我们可以设想宇宙创生之初的某些遗迹留存到今天。于是，寻找宇宙创生的遗迹现在就成了最热门的科研工作之一，而且，或许人们似乎不敢相信，这工作竟很有经济意义！原初的宇宙为我们提供了一个天然实验室，地球上最复杂的科学设备都不能模拟出来的超高温超高压的物理条件，在这天然实验室里却得到实现。为了验证他们关于物质在那些极端条件下的行为的理论，物理学家们必须求助于有关宇宙初创时的知识。他们希望，宇宙今天仍可能保留下一些遗迹以给我们提供一些线索，使我们得以了解有关宇宙创生之初短短的一瞬间所发生的物理过程。然后，就可以通过计算来看看那些物理过程与理论家们所提出的关于物质在极端条件下的行为理论是否相符合。

在20世纪60年代中期，原初宇宙最重要的遗迹被偶然地发现。两位贝尔电话公司的物理学家无意中发现来自空间的某种神秘的辐射。经过仔细分析，揭示出这种将整个宇宙都浸润其中的辐射是原初宇宙高热的遗迹，是原初宇宙火球正在暗淡下去的最后光芒。宇宙大爆炸，就像任何爆炸一样，产生了大量的热。宇宙气体花了100000年才冷却到太阳表面的温度。又过去了180亿年，宇宙温度降低到谷底，只是比绝对零度高3度。然而，现在仍有大量的能量被锁闭在宇宙热辐射里。

知道了作为宇宙创生遗迹的热辐射现在的温度是多少，那么，要想得到它在各个时期的值就是个简单的计算问题了。宇宙典型的区域在体积上每扩大1倍，其温度就要下降50%。反过来计算，就很容易推算出，宇宙创生后第1秒时的温度是100亿度。这似乎是相当热了，但这个温度完全是在实验室可模拟的范围之内。运用现代粒子加速器产生高能对撞，就有可能在极短的一瞬间模拟出原初宇宙爆炸后亿分之一秒时的情况，当时的温度是令人咋舌的千万亿度。因而，天文物理学家可以相当有把握地模拟出很多个大爆炸发生之后的物理过程。

利用这样的模型，就可能通过计算得知宇宙在爆炸诞生之后的每

个时期，宇宙间的物质是什么形态。例如在大爆炸发生之后的5秒到5分钟这一期间，当时的物理条件适合发生核反应。主要的物理过程是氢核通过聚变成为氦以及一些氘。根据计算预测，氦与氢质量的最终比率应是25%，这个比率很接近我们今天所观测到的这两种元素的相对宇宙丰度（氢与氦构成了宇宙中99%以上的物质）。计算预测与实际测量的结果有如此惊人的一致，这使我们对热大爆炸理论基本概念的正确性有了信心。

大爆炸发生之后1秒之内的各个时期涉及极高能物理。在那极高的温度下，物质完全破碎，其主要构成成分（将在第十一章讨论）呈裸露状态。这极早期即宇宙诞生后的第1秒现在是理论物理学家们的热门研究课题。有一些理论物理学家认为，早期宇宙的很多情况都可以用当时发生的物理过程给以解释。在接下来的一章里，我将描述这方面某些较新近的发展。

大爆炸理论在很大程度上已被天体物理学家看成是理所当然的了，而氦丰度的计算长久以来也已经成为标准宇宙学的一部分了，因而，那些早先取得的科学研究成就的重大意义也就容易被人忽视。要是19世纪的一个考古学家声称发现了伊甸园，并拿出了一件文物可以确凿无疑地证明是上帝创世第一天的作品，那么，他的话就得引起轰动。氦或许是大多数人所不大熟悉的，但很容易买到它的工业制品。这平平常常的实验室制备的物质竟是在原初宇宙的火炉里形成的，它不是在那火炉燃着的第一天而是在那头几分钟里形成的，这真是太不可思议了。

尽管当前的科学观点为创世论提供了强有力的证据，然而，我们还得要记住，宇宙从此以后能长生不老这种看法却找不到任何逻辑的理由。从物理学上看，宇宙永生的主要困难是热力学第二定律，这一点我们已经说过。但是，物理学家们时时提出一些机制来克服这一困难，其中之一是赫尔曼·邦迪、托马斯·戈尔德、弗雷德·霍伊尔提出的稳态理论（the steady-statetheory）。这一理论有很多翻版，但所有

的翻版也都一致认为，宇宙在年龄上是无限的。提出这一理论的物理学家也明白，若说宇宙的年龄是无限的，便会碰上热力学意义上的热死问题，但他们绕过了这个问题，办法是假定新的低熵物质在被不断地创造出来。这样，物质就不是一下子出现在一场原初的大爆炸中，而是在极长的时间里逐渐地、或者很可能是零星分散地在多次小爆炸中出现的。新物质出现的平均比率是受调整的（很可能是受某种反馈机制的调整），这样，当宇宙膨胀，现存物质的密度降低时，新出现的物质就会填充空档，因而维持了大致均衡的宇宙物质密度。于是，众星系离散形成的越来越大的空地方就这样被新创生的星系填补了，宇宙的总体也就经历了多少时代而面貌不变。从整体上看，一切都没有变化（见图3）。大爆炸模型则与此相反，星系的密度在稳步地减小，宇宙的结构和排列都在演变。

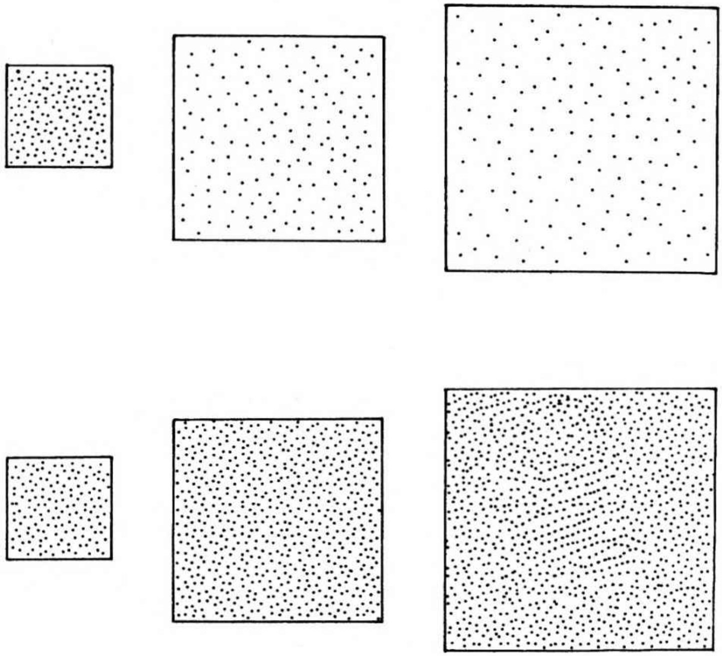


图3　有两种宇宙模型，一为大爆炸模型，一为稳态模型。图为两组“快照”，分别表现的是膨胀空间的某一区域在两种不同模型中的演化情况。在大爆炸模型中（上），星系（小点）的数目在一给定的空间体积中保持不变。这样，随着膨胀的进行，小点的密度下降。在稳态模型中（下），星系的密度在各时代都不变，于是新星系就得不断地被制造出来以便填补空间膨胀造成的空缺。

霍伊尔试着提出一种新型的、携带负能量的场来解释物质持续不断的创生。这种场的稳步增强提供了物质创生所需的正能量（能量变物质将在下一章里描述）。于是，稳态模型中就完全没有上帝的地位了。因为，第一，创造物质所必需的原初能量不必被创造，只要将负能量储入另外的某个系统，就会有原初的能量；第二，空间和时间不是由创生而来的，而是一直就存在的。

稳态模型在哲学上对很多科学家有极大的吸引力。这些科学家们

为这一模型的优美、简洁所吸引。然而，天文学的进展了结了这一理论的所有简洁版本，1965年发现的宇宙背景热辐射道道地地成了钉在装殓这一理论的棺材上的最后一颗钉子。不过，它仍然是一重要的理论，因为它证明，一个既不是突然创生、也不会有热死的宇宙在逻辑上是可能的，而且，在这样的一个宇宙中，一切过程、包括物质出现的过程，都可以在自然的机制中找到原因。

现代宇宙学为天地的创始提供了强有力的物理证据，这一事实很让宗教思想家们满意。然而，光有天地创始还不够。圣经告诉我们说，上帝创造了宇宙。科学能够对大爆炸的原因提出任何解释吗？这个问题，就是下面一章的主题。

# 第三章　上帝创造了宇宙？

我想知道上帝如何创造了这个世界。

爱因斯坦

我用不着这样的假说。

彼埃尔·拉普拉斯《致拿破仑·波拿巴》

最近，一家著名的期刊以大字标题宣布："天文学家们发现了上帝！"那篇文章的主题是大爆炸以及我们在了解宇宙极早期方面的新近进展。在通俗报刊文章里，宇宙创生的本身就被认为是足以揭示出上帝的存在的。但是，宇宙的创生难道真的是表示着上帝造成了宇宙的创生吗？能不能设想一种没有上帝的宇宙创生？难道现代物理学真是终于要揭示出物质宇宙的局限、并迫使我们去乞灵于超自然的东西吗？

创生（Creation）[2]这个词有许多意思，清楚地辨明其众多意思之间的区别是很重要的。宇宙的创生可被认为是表示突然之间，物质从那混沌的、无结构的原始形态被组织到现在我们观测到的复杂的秩序和活动中。宇宙的创生也可能真的表示在先前无形的虚空中创造出物质。或者，宇宙的创生也可表示整个的物质世界，包括空间和时间，从无中突然出现。另外，还有生命及人的创生问题，我们后面再谈。

圣经上说上帝"头一日"创造了宇宙，但没有讲明当时具体是怎么回事。现在，实际上有两种上帝造天地的说法，但两种说法都没有明确地说，用来造恒星、行星、地球以及人的身体的物质在造天地这件

事之前就存在了。基督教教义中有一历史悠久的一条，这就是相信上帝从无中创造出宇宙间的物质。倒也是，假如认定上帝无所不能，便得相信这一点。这是因为，假设上帝没有创造物质，这就是暗示，上帝在进行创造的过程中受到了原材料的性质的限制。

20世纪之前，科学家和神学家都认为，物质是不能用自然的方法创造（或毁灭）的。当然，物质的形态有变化，比如，在化学反应的过程中，物质的形态发生变化，但物质的总量被认为总是恒定的。科学家们在面对物质起源这道难题时，倾向于相信宇宙的年龄是无限的，这就避开了创造物质的问题。在永恒的宇宙中，物质可以是一直就存在的，于是，物质的起源这道难题也就被绕开了。

那种认为物质不可能以自然的方法创造出来的信念，在20世纪30年代戏剧性地崩溃了。当时，头一次在实验室里造出了物质。导致这一发现的事件便是现代物理学处于全盛时期的最好例证。

这一发现过程的故事，如同其他很多同类的故事一样，是由爱因斯坦在1905年开的头。他那有名的$E=mC^2$方程式就是“质量与能量等同”这一陈述的数学体现。质量有能量，能量有质量。质量是物质的定量，说一物体的质量是多少，就是说它有多少物质。质量大就意味着沉重，难移动；质量小就意味着轻快，容易移动。质量等同于能量这一点就意味着在某种意义上说，物质是“被锁闭的”能量。如果能找到什么方法把它释放出来，物质就会在能量的爆炸中消失。反之，如果想办法把能量集中起来，就会出现物质。

爱因斯坦的方程式是他相对论的一个副产品，在一开始进行构思时，该方程式所涉及的是以接近光速的超高速运动的物体性质。根据相对论，某物体运动所耗用的能量应该使该物体显得变重了（质量增加）。当物体以平常速度运动时，质量的增加是微小的，因为一丁点质量就相当于很大很大的能量，举例来说，1克质量相当于以目前的价格要花10亿美元才能买到的能量。然而，现代亚原子粒子加速器却

能把电子和质子的速度提高到十分接近光速，获得这样的速度之后，人们观测到电子和质子的质量增加了几十倍。

当然，质量随着速度的增加而增加还不是物质的创生，只不过是现存的物质增加了重量而已。保罗·狄拉克1930年前后进行的划时代的数学推导，使得通过聚集能量造出全新的物质粒子的可能性出现了。狄拉克当时正试图将爱因斯坦的相对论及其$E=mc^2$与20世纪物理学的另一重大革命——量子论——调和起来。量子论涉及的是原子及亚原子物质的行为。描述近于光速运动的亚原子粒子，需要有一个统一的相对论量子论。能量的放射性发射就产生接近光速运动的亚原子。

狄拉克进行了数学分析之后，提出了一个新方程式来描述高速的原子物质。该方程式立刻获得了成功，解释了一个当时一直令人困惑的电子的特征，即电子总是以与常识和基本几何完全相悖的方式旋转。粗略地说，一个电子得自旋两周才能再现出同一个面孔。狄拉克方程式又是一个好例，说明在探寻基本物理学的抽象世界时，数学如何必须取代直觉。

然而，狄拉克的方程式却还有令人不解的一面。方程式的各个解答正确地描述了普通电子的行为，但每一个解答都伴有另一个解答，这种相伴解答在已知的宇宙里似乎没有任何对应物。只要稍微发挥一下想象力，就可能想出这未知的粒子是怎样的东西。它们的质量和旋转方式与普通的电子一样，只不过所有的电子都带负电荷，而这些神秘粒子却要带正电荷。它们的其他特征，如它们的旋转，也要与电子相反，使这种新粒子成为电子的一种镜像。

更有意思的是狄拉克的预言。他预言说，假如能聚集起足够的能量，以前没有这种粒子的地方就会出现这么一个“反电子”。为了电荷守恒，出现这么一个“反电子”的同时，必定也得出现一个电子。这样，能量就可以以电子——反电子的形式直接用来创造物质了。

大约在这个时候（1930年），物理学家赵忠尧正用铅之类的重材料做γ射线（高能光子）穿透力的试验。他注意到，最有力的γ射线被以一种无法理解的高效率减弱了。γ射线被额外吸收，其原因当时对赵来说是个谜。但我们现在知道，那是因为生产电子一反电子的缘故。

后来，到了1933年，卡尔·安德森在用金属薄板研究宇宙线（即来自宇宙空间的高能粒子）时，头一次明确地发现了狄拉克所预言的反电子。在实验室里的受控实验中，物质被造出来了。接着很快证实了，新发现的粒子具有已预言过的一切特征，于是，狄拉克和安德森因这精彩的预言和发现分享了一项诺贝尔奖。

在其后的年月里，制造电子和反电子（反电子通常被称为正电子）成了范围广泛的实验室实验过程中的家常便饭。第二次世界大战之后，亚核粒子加速器的发展也使得受控生产其他种类的粒子成为可能。反质子、反中子也制造出来了。今天，正电子和反质子可以大批量制造，并被储存在磁“瓶”里。镜像粒子或曰反粒子都被称为反物质。现在，人们在物理实验室里例行公事般地制造反物质。

有了这些事实做后盾，人们似乎摸到了门路，可以对一切物质的本源给出一个自然的解释了。发生大爆炸时，有大量的能量，使物质和反物质得以断续地大量产生出来。最后，爆炸热冷却下来，这些物质就聚集成恒星和行星。不幸，这简洁干净的解释却有个大漏洞；因为物质遇上反物质时，两者会彼此湮灭，同时会有剧烈的能量释放，而这正与物质创造的过程相反（见图4）。

因而，由物质和反物质混合组成的宇宙是极不稳定的。在我们的星系中掺加反物质得有非常严格的条件限制，因而，反物质的量是微乎其微的。

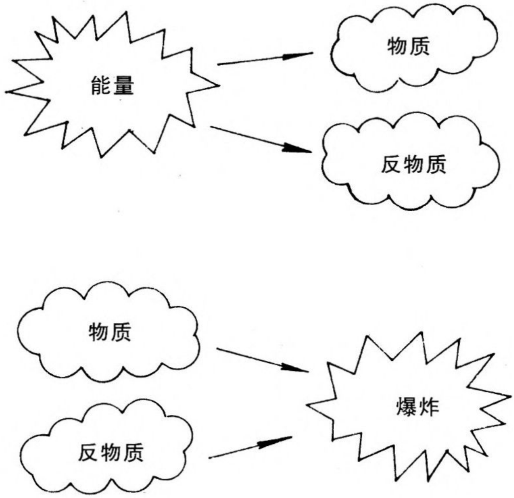



图4　在实验室中，可以用能量创造物质，但是，物质总是与等量的反物质一起出现的。当物质和反物质相遇时，就会发生爆炸，使双方湮灭。同时放出锁闭在物质中的能量。人们不知道，宇宙间的一切物质是怎么被创造出来，同时又没有掺上可以使其爆炸湮灭的反物质。

那么，那些该有却不在的反物质都到哪里去了呢？在实验室中，每造出一个粒子，就必有一反粒子。照理说，宇宙中的物质与反物质之比该是一比一，但多次观测的结果却排除了这照理说是该有的事。有些天体物理学家提出一种假说，认为物质和反物质不知怎么彼此分离开来，彼此天各一方了。这些天体物理学家试图以这种假设来解释反物质的失踪之谜。或许，一些星系是由反物质组成的，另一些星系是由物质组成的。然而，迄今为止，还没有人提出过令人信服的物质与反物质的分离机制。因而，对称宇宙论失宠了。

于是，那些坚持认为大爆炸与宇宙创生是一回事的科学家们，显然就必得假定，某一超自然的过程以违反一切物理学定律的方式，把没有反物质相伴随的物质注入进宇宙。然而，“在奇点处一切定律都失去了效用”这一类的含糊其辞，并不能减轻人们的疑虑。

但最近，出现了一条或许能走出这一困境的路径。尽管在实验室条件下，物质和反物质总是成双产生的，但在大爆炸的超高温条件下，很可能可以多出一点点物质。那种试图对自然界的四种基本力给以一个统一的描述的理论研究，使人们产生了上述的想法（在第十一章里，还要对此进行较为详尽的讨论）。根据理论计算，在温度高达$10^{27}$度时（这一温度只能在宇宙创生的头$10^{-36}$秒才能达到），每产生10亿个反质子，同时就会产生10亿零1个质子。同理，电子也会比正电子多出10亿分之1。

虽然只多出这么一小点，但其意义却十分重大。在后续的大厮杀中，10亿成对的质子和反质子彼此湮灭，留下了一个未配对的质子和一个孤立的电子。这些近乎是大自然计划外的添加物的剩余粒子，就变成了后来构成所有星系的物质，所有的恒星和行星，还有我们人类本身，都是由这些物质构成的。根据这一理论，我们的宇宙是由微量剩余的非均衡物质构成的，这些物质是那难以想象的瞬间大爆炸的残留物。

如同一切好理论一样，物理学家们觉得，这样解释物质的起源有相当的说服力。但到哪里去找硬邦邦的证据呢？

现在似乎有了两个可验证的计算结果。第一个结果与大爆炸之初的10亿粒子与反粒子大规模互相湮灭有关。10亿粒子和反粒子相互湮灭之后，还会剩下一个多余的粒子，同时，伴随湮灭所释放出的能量也必会留存下来，很可能就是以热的形式留存下来。正如我们在前面的一章里讲到的那样，宇宙确实是浸泡在大爆炸所留下的热辐射里。那么，把每一现存的原子的热能加起来，看看其数目是否与十亿分之一的计算相符合，这是很简单的事。结果发现数字的确相符，至少可以用很说得通的模型来找出一致性。因此，这一理论不但解释了物质的起源，而且也说明了宇宙的确切温度。这确实是项了不起的成就。

然而，还需要更进一步的证实，才能够有把握地宣布物质的创造

不是来自神。假如实验室能够提供某种直接的证据，证明物质与反物质之间确有明显的不对称，那将会最让人信服。我们很可能快要幸运地获得这样的证据了。

预言物质产生会有微量剩余的理论，也同样预言：根据产生微量剩余物质的机制，也会有物质的自发性微量毁灭。这种理论认为，在极长的时间里，质子会衰变成正电子，而正电子会进而湮灭电子。这样，一切物质最后注定要归于消失。但其时间尺度是如此之长，以致一个人的身体在其一生中平均才失掉大约一个质子。为了验证这一理论，科学家们正在地下深处（为的是消除宇宙线的干扰）研究观察极大量的物质，试图揪住一个正在消失的质子。因为这样的研究观察过程从根本上说是统计性的，所以要耐心地等上几个星期才能观测到偶然的反常衰变，要知道，一个质子的平均寿命至少是$10^{30}$年。这研究观察过程的原理就是堆积起成千上万吨的物质（这些物质就代表着大量的质子）以找出偶然的随机事件。现在，正在进行着好几个这样的试验，而其中至少一个实验已经显示出一些可能的质子衰变事件。

任何人若想从物理现象中推导出上帝的存在，都得面临一个基本的问题，这问题就是物质的起源。物质在没有反物质的情况下出现，这在先前似乎是奇迹，现在却根据进步了的科学知识，似乎用普通的物理道理就可以解释了。不管某一具体事件是多么令人惊讶，多么不可解释，谁也不敢绝对有把握地说，在将来某一遥远的时刻，人们找不到一个自然现象来解释那令人惊讶的事件。

那么，上述的那些科学进展真正意味着我们可以用自然过程来解释宇宙的创生了吗？对于这个问题，很多神学家持强烈的否定态度。我们所描述的那些过程，并非从无中创造出了物质，而是将先前存在的能量转化成了物质。我们仍需要首先解释那用于创造物质的能量是从哪里来的。这难道真是得用超自然的解释吗？

然而，当讨论从物质转向能量的时候，人们得小心谨慎。能量是

个相当难把握的概念，尤其是在现代物理学中。什么是能量？能量可以有很多不同的形式，例如很可能只是运动。在实验室里，粒子可以以极高的速度碰撞，于是，以前只有两个粒子的地方，会出现四个粒子。新粒子的出现，是以那两个原先的粒子的速度降低为代价的。把那无形的运动转化成有形的物质，就很有从无中创造有的味道。

还有一种可能性更令人惊讶，这就是从能量为零的状态中创造物质。之所以会出现这种可能性，是因为能量既可以为正，也可以为负。运动的能量和质量的能量总是正的，但引力的能量，如某些引力场或电磁场的引力是负的。有时会出现这样的情况，创造新生物质粒子质量的正能量正好被引力或电磁力的负能量抵消了。例如，一个原子核附近的电场很强。假如能够造出一个含有200个质子的原子（这是可能的，但很困难），整个系统就会变得不稳定。这时即使没有任何能量输入，也会生出电子正电子对，这是因为，新生的粒子对所发出的负电能可以恰好抵消其质量中含有的能量。

引力场的情况就更奇特了，因为引力场只不过是空间弯曲，是弯曲的空间，锁闭在空间弯曲中的能量可被转化成物质和反物质的粒子。这种情况现在在黑洞附近就有，而且很可能是大爆炸时粒子的最重要来源。这样，物质就自发地从空空如也的空间里出现了。于是问题就来了，到底是原初大爆炸具有能量？还是整个宇宙是一种能量为零的状态，其中一切物质的能量已被引力场引力的负能量所抵消？

简单地计算一下可能就能解决这个问题。天文学家们能够测量诸星系的质量，星系间的平均距离.以及它们退行的速度。把这些数字代进一个公式，就能得出一个数字，而某些物理学家已经把这个数字解释成宇宙的总能量了。这个数字在可观测的精度里的确是零。为什么会有这样的结果？宇宙学家们长久以来一直迷惑不解。有些宇宙学家提出，有一个深藏不露的宇宙原理在起着作用，根据这一原理，宇宙的能量就得恰好为零。假如真是这样，宇宙就可以走那阻力最小的路，用不着输入任何物质或能量就可以诞生了。

有一事实使问题更复杂化了。这事实是：当引力存在时，能量是不能确切地规定的。在某些情况下，可以把某一封闭系统所受到的引力影响看成是无限远，以此来确定该系统中的总能量。但在一个空间有限的宇宙中，如我们在上一章简短讨论过的爱因斯坦所提出的宇宙模型中，这种办法就完全失效了。在这样一个封闭的宇宙中，宇宙的总能量是个无意义的量。

上面所讲的从空空如也的空间中或许连能量也不用输入就自然产生物质，这一类的例子真能使人们不用神学也能解释宇宙的创生吗？有人或许会说，科学仍没有解释空间（及时间）的存在。物质的创造长久以来一直被认为是神的行为结果。即使现在可以用普通的科学知识来解释物质的创造了，但是，为什么会有宇宙，为什么首先得有时间和空间，然后物质才能从中产生出来，解释这类问题难道只能求助于上帝吗？

那种认为宇宙作为一个整体有一个原因，而这原因就是神的看法，是由柏拉图和亚里士多德提出来的，经过托马斯·阿奎那的发展，在18世纪由莱布尼茨和塞缪尔·克拉克使之臻于完善。这种看法，通常称作上帝存在的宇宙论论证。这宇宙论论证有两个变体，一个是我们将在本章里讨论的因果论证，一个是将在下一章里讨论的从偶然性出发的论证。大卫·休谟和伊曼纽尔·康德对宇宙论的论证持怀疑态度，而贝特兰·罗素则对之进行了严厉的攻击。

宇宙论论证的目标有两个。一个是确立“第一推动者”的存在，有了第一推动者就可以解释世界的存在。另一个目标是证明这第一推动者就是通常的基督教教义中的上帝。

宇宙论论证的推理过程是这样的：毫无疑问，每一事件都必有一个原因，但原因的链条不可能是无限的，所以万物必有一个第一因。这第一因就是上帝。必须说明的是，宇宙论论证有很多变体，而且其意义也有很多微妙的解释，因此，多年来关于宇宙论论证的辩论已经

变得繁复难解了。在这里，我不想对辩论的各方进行一番公允的评价，我只是想说，宇宙论的论证受到了人类历史上一些最伟大的天才们的注意，但这并没有使辩论的双方避免犯一些逻辑上和哲学上的大错。我们这里要做的，是用现代科学的观点重新研究因果链假说。

我们不妨研究一下宇宙论论证的第一步：每一事件都有一个原因。克拉克曾声言："最令人百思不解的是，世上竟然有东西，而且与没东西相比，有东西是毫无道理的。"[①]可以说，人们通常都是假定，每件事的发生都是由另一件事所造成的，一切进入存在的东西都是由一些早已存在的东西造成的。这似乎很有道理，但事情果真如此吗？

在日常生活中，我们很少怀疑一切事件都有原因。一座桥坍塌了，那是因为它负荷过重了；雪融化了，那是因为空气热起来了，一棵树长出来了，那是因为种下了树种；等等，不一而足。但是，会不会有些东西没有原因？

考虑考虑上面的断言吧："一切进入存在的东西都是由某种东西造成的。"假如某种东西从来没有进入存在，而是一直存在着呢？这样的东西确实可以想得到，例如，稳态宇宙中的空间就是。一个永恒存在的东西、一个从来没有不存在的东西是否有一个原因，这样的问题难道有什么意义吗？人们可以追问："为什么这东西存在？"若是回答："它一直存在。"这样的回答未免有点蹩脚。因为人们完全可以设想这东西可能不存在，所以不管其年龄多长，询问它为什么存在而不是不存在似乎也是很合理的。因而，照某些人看来，取消创生（如稳态的宇宙就用不着创生）并不能排除解释宇宙为什么存在的必要性。

我们现在暂且放下永恒存在的东西，专门考虑一下东两开始存在的情况吧。某个东西能从无中产生吗？我们说过如何从空空如也的空间产生出粒子，但是，那是因为有空间弯曲。要想说出粒子的起源，我们就得说出空间是从哪里来的（假如空间不是一直存在的话）。有

些人或许会提出疑问：空间算不算一种东西？不错，很难想象托马斯·阿奎那或莱布尼茨把空间看作是因果链的一部分。不过，我们继续往下想吧。是什么使得空间在大爆炸中突然出现？是奇点？但奇点肯定不是一种东西。奇点只是一种东西（时空）的边界。此路不通。

每一事件都有一个原因吗？某种东西在没有任何在先的作用力或任何合理的理由的情况下能够出现吗？报纸上常常大肆宣扬“天空中发现无法解释的物体。”然而，这并不是说那些空中的现象没有解释，而只是说目前没有已知的解释。不幸的是，目前难以看到“每一事件都有一原因”这个断言一朝会被确凿地证明是假的，因为若想证明其为假，就不但得找出一个没有原因的事件，而且还得进而证明，不管人们对宇宙的了解有多详尽，不管人们对自然界的知识有多深入，仍是不可能找到任何原因的。这似乎是不可能做到的事。谁能有把握地说，那令人迷惑不解的事件不是由某个完全隐蔽的、极其罕见的、从未出现过的、难以察觉的偶然过程造成的？

在证伪“每一个事件都有一原因”这一说法方面，进展最大的学科是量子力学。我们在第八章就可以看到，在亚核世界里，粒子的行为通常是不可预测的。你不可能确切知道，一个粒子在这一时刻和下一时刻之间要干什么。假如人们要选择一亚核粒子到达某一具体位置作为一个事件，那么，根据量子论，该事件就是无原因的，意思是说，它本质上是不可预测的。不管我们对作用于该粒子的所有的力以及所有的影响了解多深，我们仍无法说该粒子到达规定的位置是由某种其他的东西“造成”的。该粒子的运行轨迹本质上是随机的。它只是毫无节奏又毫无理由地突然在那里蹦出来。

某些（少数）物理学家对量子论的这一观点不以为然。爱因斯坦用一句名言反驳它道：“上帝不掷骰子。”在这些物理学家的心目中，每一事件都应该是由这种事或那种事造成的。即便是在亚核层面上也是如此。令人吃惊的是，现在已能够进行一项试验，证明原子系统在本质上的确是不可预测的，除非影响的运行速度超过光速。“上帝”确

实掷骰子。只要大自然没有玩弄异常的把戏来搞乱这些试验结果，那么，上帝掷骰子的说法就近乎有相当坚实的理由。

因而，假如人们承认，量子事件个别地来看是没有直接的原因的，那么，物质创生这一量子过程的经典例子也能说是没有物理原因的吗？从某种意义上说，是没有物理原因。一个单个的粒子会以不可预测的方式突然出现，你没法知道它在哪个时刻、哪个位置出现。然而，该粒子的行为尽管放荡不羁，却仍然受制于概率的规则。假如有某一程度的空间弯曲，那么某个粒子就很可能在某一时间出现在某一空间区域。但谁也说不准它什么时候、在哪里出现。反过来说，那个粒子也有确定的机会现在就在你的起居室里突然蹦出来，尽管这几率非常之小。在量子世界里，这种事随时都有发生。粒子创生的几率决定于空间弯曲的度数，这一事实就意味着某种大致的因果关系。空间弯曲使得粒子的出现更为可能了。是否应严格地把这看成是粒子出现的原因，这在很大程度上是个语义学问题。

有人或许会提出异议，说我们讨论的中心问题是“整个宇宙是否有个原因”，而不是电子的创生或电子到达某一位置是否有个原因。某些物理学家无疑会回答说，整个宇宙也是受量子原理的制约的。但这么一来，我们就得扎进那令人心烦的量子宇宙学，而这门学科本身也有一大堆自恰的问题（在第十六章里才能进一步讨论这一问题，我将提出一个量子方案，或许可以解出宇宙起源的难题）。现在，我们暂且撇开量子论，承认整个宇宙可以说有一个原因，那么，其原因是什么呢？是上帝吗？

现在，我们往下来研究一下宇宙论论证的第二步：不可能有一个无限长的因果链。链条总有个头。星系是由旋转的星系形成的，星系是由原初的氢气形成的，氢是由大爆炸的一瞬间产生的质子形成的，质子是由空间弯曲形成的。人们总是认定这因果序列一定有第一项。阿奎那写道：

在可观测的世界中，人们发现因果是排列有序的；我们从未观察到，也永远不会观察到某种东西是它自己的原因，因为真是它自己的原因的话，这就是说它先于它自己，而这是不可能的。然而，因果序列必有一尽头，因为在一因果序列中，前一项是中间项的原因，中间项是末项的原因（中间项是一项还是多项无所谓）。若消除一个原因，这原因引起的结果也就消除了。因此，假如没有第一项原因，就不可能有末项原因，也就不可能有中间项原因。假如该因果序列是没有尽头的，因而也没有第一项原因，那么，也就不会有中间项原因，不会有最后的结果，而这显然是错误的。因而，人们就只有设想有个第一因，大家都把它称作“上帝”。②

在争辩说因果链不可能没有尽头的过程中，阿奎那和克拉克都没有以因果链本身是无限的为理由进行反驳。这两位思想家提出他们的论点的背景是，当时的人们认为宇宙是永恒的，年龄是无限的，大家都愿意把宇宙创生的证据寄托在“神的启示”上，而不愿为此进行理性的探讨。相反，他们提出的异议似乎是，一条把整个宇宙也贯穿其中的无限长的因果链据信是不可能的：

假如我们思考一下一个这样的无穷序列……显然，这整个一系列的实在物不可能有外部的原因；因为这序列据认为是包括了宇宙中一切现有的和已有过的东西；再者，这序列的存在显然不可能有内部的原因，因为这无限序列中的每个实在物都不能被认为是独立存在的或必然的……而是每个实在物都要依存于前一实在物……因而，一个相互依赖的实在物的无限序列，若没有任何原初的独立原因，它就没有必然性，也没有有内部或外部原因的一系列实在物：这就是说，它是一个显然的矛盾体，是不可能的事。③

由相互依赖的实在物构成的无限序列，不那么严格地说就是一个无限的因果链，它的存在需要有个解释（这因果链包括了一切存在的东西，因而就找不到解释），这种看法，受到了哲学家们的激烈攻击，尤其是休谟和罗素。罗素与柯普莱斯顿神父在英国广播公司电台

上进行了一场有名的辩论。罗素这样陈述他的论点：“每个存在的人都有一个母亲……但人类显然没有一个母亲。”简单地说，他认为只要解释了序列中的每个个别成员，那么，根据事实本身，也就解释了整个序列。因为因果链的任何一环的存在，都依存于某个或某些前面的链，这样，无限因果链的每一环都得到了解释。要求说出整个宇宙的一个原因与要求说出宇宙中某一个别物体或事件的原因，这两者在逻辑上不是一回事。

事实上，“集合的集合”这个题目难对付是有名的。假如一集合被无害地定义为任何（具体或抽象的）事物的集，那么，就像罗素用他那有名的悖论所表明的，一个由集合构成的集合根本就算不上是个集合！比如，我们可以把某一图书馆的书的总目录看作是一个集合，但是，目录本身要不要列入目录？有时候要。我们把这种目录本身也列入目录的目录，叫作“TypeI”，把目录本身不列入目录的目录叫作“TypeII”。我们再把存在中心图书馆的主目录看作由集合构成的一个集合。这主目录的功用是，列出一切TypeII目录，主目录是目录的集合。这够合理的吧？不幸得很，那由所有TypeII目录组成的集合是自相矛盾的。我们只要问一问“主目录本身是TypeI还是TypeII，”我们马上就可以发现其自相矛盾。假如主目录是TypeII，那它就不包括它自己。但主目录按其定义是列出了一切不包括自己（TypeII）的目录的。因此，主目录包括它自己，因而它是Type工。但这也不可能，因为主目录只是列出TypeII目录，因而，假如主目录是TypeI它就不能列上它自己。因而，它没有列上它自己，它是TypeII，结果：自相矛盾的废话。

上述的要点是，由现存事物构成的整个宇宙这一概念是微妙的。目前不清楚宇宙是不是个事物，假如把它定义为事物的集合，就不会有悖论的危险。这一类的问题布下的陷阱，等待着所有试图从逻辑上证明上帝的存在是一切事物的原因的人。

即便承认宇宙论论证，承认宇宙肯定有一个原因，但把原因说成

是上帝却仍有一个逻辑困难，因为人们接着可以问：“什么是上帝的原因？”对这样的问题，通常的回答是：“上帝不需要原因。上帝是必然的存在，其原因要在其内部寻找。”宇宙论论证的基础，就是设定一切事物都必有一原因，然而其结论则是至少有一个事物（上帝）不需要原因。因此，宇宙论论证似乎是自相矛盾的。而且，假如人们准备承认某个事物（上帝）可以在没有外在原因的情况下存在，那还有什么必要去唠叨那无穷无尽的因果链？宇宙为什么就不能在没有外在原因的情况下存在？设想宇宙是其自身的原因比起设想上帝是其自身的原因来，难道还需要更大的巧思吗？

假如我们停下来，不再往前走（走得比上帝还远），那又何必走到这一步呢？为什么不在物质世界处停下来？……我们设想它在其自身内部包含了它的秩序原理，我们实际上就是断言它是上帝。④

休谟的这段话使人想起许多科学家所持有的模糊信念：“上帝就是大自然”或“上帝就是宇宙。”

或许，对因果变体的宇宙论论证最厉害的反驳是这一事实，即因与果是牢固地嵌入时间概念中的概念。然而，正如我们已经看到的那样，现代宇宙学提出，宇宙的出现牵涉到时间本身的出现。通常公认，原因在时间上总是先于结果。例如，放了枪之后，靶标才破碎。这样，按照通常的因果意义谈论上帝创造宇宙显然是无意义的。假如创造宇宙的行为牵涉到创造时间本身的话，如果没有“以前”，大爆炸就不可能有自然的或超自然的原因（通常意义上的）。

圣奥古斯丁（354～430）似乎很明白这一点。有人认为，上帝等待了一段无限长的时间，后来挑了个吉祥的时刻创造了宇宙。圣奥古斯丁嘲笑了这种观点。他写道：“世界与时间都有一个开始。世界并不是在时间中被创造出来的，而是与时间一起被创造出来。”⑤考虑到在奥古斯丁的时代，人们对时间空间的各种错误看法，那么，奥古斯丁的确了不起，他预言了现代宇宙学的观点。

奇怪的是，奥古斯丁对创世纪的深刻见解，在13世纪基督教会接受了古希腊传统的影响之后，却受到了挑战。在后来的争议中，第四次拉特兰会议（1215）反驳了亚里士多德关于宇宙年龄无限的哲学，认定宇宙在时间上确有开始，并以此作为基督教的一个信条。但即使在今天，神学家们对圣经的创世纪仍是见解不一。

有一个超越时间的上帝存在的假定所带来的问题是，尽管这么假定会使上帝来到“此地此时”，但大部分人认为是上帝所具有的很多美质只有在时间的框架中才有意义。难道上帝不能作计划，应答祷告，对人类行进的路程表示喜悦或担忧，然后再进行审判吗？难道他不是不停地在世界上活动，做工作，“给宇宙机器的齿轮上油”，以及干诸如此类的事吗？上帝的这一切活动，若不是在时间的构架里就全无意义。若是不在时间之中，上帝怎么能计划、行动？假如上帝真是超越了时间，因此预知了未来，他为什么还关注人类的进步或关注与罪恶的斗争？结果他早就知道了的（我们将在第九章里再来讨论这个问题）。

事实上，正如我们已经看到的那样，上帝创造了宇宙这一观念本身，就指的是一个发生在时间之中的行动。我上课讲宇宙学时，常有人问我大爆炸之前发生了什么事。我的回答是，没有大爆炸的“之前”，因为大爆炸就代表了时间本身的出现。我的回答受到怀疑——“肯定是什么事引起了大爆炸。”但因果是时间概念，是不能应用于其中不存在时间的状态的。因而，大爆炸之前发生了什么事之类的问题是没有意义的。

假如时间的确有一个开始，那么，谁要想用原因来解释时间的开始，就必须求助于一种更广泛的原因概念，这概念与我们日常生活中所熟悉的原因概念是不同的。一个可能就是不再要求原因总是先于结果。原因能够逆时间而动，造出先于它的结果吗？当然，改变过去，这种概念是充满了悖论的。比如，设想你能够影响19世纪的事件来阻止你自己的出生，这岂不荒唐？然而，现代物理学确有若干理论涉及

逆动因果关系。假设的超过光速的粒子（叫“速子”）就能有这种逆动的因果。为了避免悖论，人们可以设想原因和结果之间的联系是非常松的，而且是不可控制的，或者，其联系是难以捉摸的。我们就要看到，量子论需要一种反过来的时间因果关系。比如，今天进行的一项观测可以影响到遥远的过去的实在建构。约翰·惠勒曾对此强调说：“量子原理表明，从某种意义上看，一个观测者在将来做的事就决定了过去发生的事，甚至那过去是如此遥远，远到连生命都不存在。”⑥

这里，就像人们在量子论中不得不做的那样，惠勒给了精神（“观测者”）一种重要意义，因而涉及了宇宙演化后期出现的精神与宇宙的创生的关系问题：

假如宇宙在其未来的历史中，不能保证在某处及某一小段时间里产生生命、意识以及观测者，那么，宇宙诞生的机制岂不是无意义或不可探测、或既无意义又不可探测的吗？⑦

惠勒希望，我们可在物理学的构架中，发现一个使宇宙能“自动”诞生的原理。在探索这样的理论的过程中，他说：“任何指导性原理似乎都不如要原理为宇宙诞生提供一种方式这种要求更为有力。”⑧惠勒把这种“自发”宇宙比作电子学中的自激电路。

即便是能从某种晚近的自然能动性（不管它是精神还是物质）里发现时空创生的原因，也难以明白宇宙怎么能自然地从无中创造出来。仍然得要有一些“原材料”，来供精神或其他什么东西进一步进行追溯性工作。惠勒指出，空间和时间都是合成的结构，就是说，二者都是由他叫作前几何（pregeome-try）的构件构成的。其他许多物理学家也指出，空间与时间不是基本的概念，而是些近似值。正如表面上看去是连续的物质实际是由原子构成的那样，时空或许也可能是由一些更基本、更抽象的实体构成的。这可能是试图发现引力的量子论的一个结果（引力只不过是时空几何而已）。在某些极端的物理条件

下，假如在大爆炸之初，时空可能“分离开来”，从而暴露出其内部构件。用时间之前的语言说就是，大爆炸可能就是一个事件，在该事件中，那些“齿轮”紧密地结合起来组成一种外表看来是连续的时空。根据这种观点，大爆炸是空间、时间及物质的开端，但不是物理学的界限。在大爆炸之外（不是在其“之前”，因为没有之前），存在着一些没有装配起来的“齿轮”，一些形而下的东西，但不在时间或空间中。

宇宙如何诞生，询问宇宙是否有某种原因这类问题是否有意义，在这些话题结束之前，我们必须考虑回答是“是”的可能性，但宇宙的某种原因不一定是上帝。正如我们早先说过的，宇宙论论证的第二部分是想论证宇宙的创造者就是上帝。但现代物理学却发现了两个新的可能性，这是宇宙论论证辩护者从未能想到的。

前一章解释过，物质的创生如何用膨胀的空间（空间弯曲）就可以予以充分的说明。而且，空间伸缩力现在似乎不存在界限。最微小的区域能无限地膨胀。在宇宙创生后的十亿分之一秒，我们目前所观测到的宇宙（共$10^{27}$立方光年）蜷缩在大约太阳系大小的体积里。在更早些的时刻里，宇宙的体积更小。因而，空间可能是从无中产生出来的，物质则可能是从空间产生出来的。但人们觉得，肯定是个什么东西使无限小的微粒一样的宇宙开始爆炸性地膨胀开来，而这就是我们追寻奇点、因果关系等等的所在。

然而，还有另一种解释，可以解释我们由空间和物质构成的宇宙。可以粗略地称之为“能繁殖的宇宙”。这里最好用类比来描述。因为空间是有伸缩性的，我们就不妨设想它是块胶皮（胶皮只是二维的，而空间则是三维的。这是个概念性缺憾，但在逻辑上没有问题。我们即将描述的东西也适用于三维，但不能在三维里形象化）。

图5表示了一系列步骤。首先，胶皮上有块凸起。接着，凸起膨胀起来，同时，一直有一狭窄的“颈部”使凸起与胶皮相连。凸起变成气球状。现在，设想与胶皮相连的颈部收口，最后，口收到一起，使

气球完全封闭起来。最后，颈部断开，气球与胶皮分离开来，使颈部再次恢复成一块连续的胶皮。这胶皮实际上生出了一块完全独立的、与自己不相连的胶皮（气球），而这气球然后又可以无限膨胀。需要的话，这新生的气球本身也可以用来生出其他的气球。

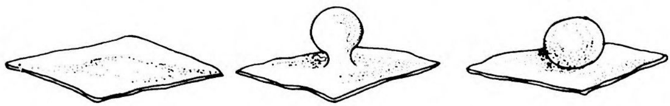


图 5　爱因斯坦的广义相对论所提出的空间伸缩性，允许“子宇宙”（气球）从“亲宇宙”（胶皮）中生长并分离出来。这种拓扑变化是在近来的某些理论中被提出来的，但没有得到人们很好的理解。

假如我们将我们的宇宙——我们用物理方式所能达到的一切空间——看成是那个“新生的气球”，那么，可以肯定，我们这宇宙并不是一直就存在的：它是被创造的。然而，其创造者仍然可在自然的物理过程中找到，即创造者是一种创造机制，该机制的本源在“母胶皮”中。那胶皮我们也不是完全不能理解，只不过它是在我们的时空之外，因此我们不能在我们的宇宙之中找到它存在的原因，但这里面牵涉不到上帝。

从这一理论中生发出来的中心问题是，那通常被看作是“宇宙”的东西，事实上可能只是一不连续的时空片断。可能有很多、甚至是无限多的其他宇宙，但它们在物理上都是其他宇宙所接近不了的。假如这么定义“宇宙”，那么，对我们的宇宙的解释就不在我们的宇宙之中，而是在其之外。但这么解释宇宙不涉及上帝，只涉及时空和某个怪异的物理机制。

最近在若干理论研究中，提出了这样一个机制⑨。在极热的条件下，可以想象空间会变得不稳定，从而以这样的方式“生出”其他的“气球”。人们甚至可以想象，有一个有着足够发达技术的集体，有意地

操纵新宇宙的创生。然而，理论纯正癖者无疑会提出异议，说关于宇宙创生的这种假说只是一个假解释，因为它仍是没有解释“胶皮与气球”的总体。说得不错，是这么回事。但上面的“胶皮”例子的确说明，原则上我们在宇宙中所能察觉到的一切事物，可能仍是由自然的原因在一段有限的时间以前造成的，在我们时空之外的东西（假如真有的话）不一定就完全是超自然的。

以上的这一通分析，对我们寻找造物者上帝又有些什么帮助呢？假如我们坚持那种简单的原因概念，不管宇宙的年龄是否是无限的，不管宇宙在时间上是否有个确定的起点，在这种情况下，那种认为万物都有一个第一原因的论点是容易受到严肃的怀疑的。奇异的机制，诸如逆动因果性或量子精神过程，可能会消除宇宙创生的首要原因的必要性。但人们仍然觉得找不到第一原因就不舒服。神学家理查德·斯温伯恩写道：

假如宇宙的年龄是无限的，因此每一时刻宇宙的每一状况都可用宇宙先前的状况以及自然法则予以完整的解释（因而这就用不着上帝了）；在无限的时间里，宇宙的存在就有了完整的解释甚至是完全的解释；谁要是这么想，那可就错了。宇宙没有解释，什么解释也没有。它是完全不可解释的。⑩

为了说明这一点，我们可以设想马是一直就有的：每一匹马存在的原因可以由其亲马的存在来解释。但我们并没有解释为什么会有马为什么世上会有马，而不是没有马，或为什么有马而没有比如说独角兽。尽管我们或许能找出每一事件的原因（这从量子效应来看是不可能的），我们仍是不能解开这个谜：为什么宇宙有现在的特征，或说，为什么会有宇宙？

# 第四章　为什么会有个宇宙？

为什么某个事物存在而不是不存在，这在大自然中有个原因。

莱布尼茨

宇宙越是好像可以理解，它也就越是好像没有意义。

斯蒂芬·温伯格

造物主上帝按他自己的意志使宇宙诞生了，这种观念在基督教-犹太教文化中有着坚实的根基。然而，我们已经看到，这种假设引起的问题比解决的问题还多，而且好多个世纪以来，它受到了神学家们的严肃疑问。问题出在时间的性质上。今天，我们知道，时间与空间是不可分割的。时空正如物质一样，也是自然的宇宙的一部分。我们将在第九章看到，时间有自己的变化和行为规律，时间明显的是属于物理学的。假如时间属于自然的宇宙，而且受物理学规律的支配，那么，时间就必然被包括在据认为是上帝创造的宇宙中。按我们通常对因果关系的理解，原因在时间上必定先于其结果。那么，说上帝是时间存在的原因又是什么意思呢？因果关系是一种时间性的活动。在任何事物成为某种原因的结果之前，必得先有时间。假如时间不存在，假如没有“之前”，认为上帝在宇宙创生之前就存在，这种天真的想法显然就是荒谬的。

我们已经讲过，这些问题，圣奥古斯丁早在5世纪就看出来了。一个世纪之后，鲍依修斯专门说明了这些问题，并把它们发挥成一个“创生”的概念。这概念比起那个大多数门外汉仍然熟悉的概念抽象得多，也难懂得多。根据他那种精致的观点，上帝是完全存在于时间

和空间之外的；从某种意义上说，上帝是“超乎”自然，而不是先于自然的。时间之外的上帝这个概念不好理解，因此，我在这里先不详细讨论它。我要把这个问题留到第九章，那时我们再深入探讨时间的性质。

时间之外的上帝被认为是以“在每个时刻都把握着宇宙的存在”这种更为强有力的方式“创造”了宇宙。时间之外的上帝并不是简单地启动了宇宙（上帝启动了宇宙这种观点被认为是自然神论而不是有神论），而是一直在行动。于是，那遥远的宇宙创造者便被赋予了一种更强的直接感——他正在此地此时活动着——但是，这样一来，他的形象也就有些朦胧了，因为超乎时间的上帝这一概念是难解的。

一种上帝是时间中的上帝，他创造了宇宙，是宇宙创生的原因。另一种上帝是超乎时间的上帝，他把持着宇宙（包括时间）的存在。这两种上帝有时人们用下面的图解来描述。①设有一系列事件，每一事件都依存于前一事件。这一系列事件可以表示为系列……$E_3$，$E_2$，$E_1$，按时间往后排。因而，$E_1$的原因是$E_2$，$E_2$的原因是$E_3$，依此类推。这一因果链可以表示如下：

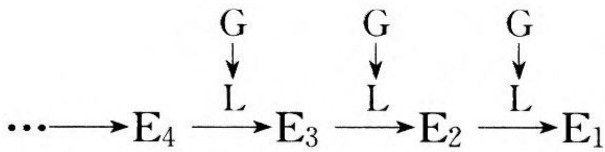
$$E_4 \xrightarrow{L} E_3 \xrightarrow{L} E_2 \xrightarrow{L} E_1$$

上面的“L”表示，通过物理定律L的作用，一事件引发了后面的事件。

上帝是宇宙的原因这一概念（即我们在上一章里详细讨论过的），可以将上帝（用G来代表）表示为这原因系列的第一项：

$$G \longrightarrow \cdots \longrightarrow E_4 \xrightarrow{L} E_3 \xrightarrow{L} E_2 \xrightarrow{L} E_1$$

另一方面，假如上帝是超乎时间的，那他就不可能属于这种因果链。他是超乎因果链的，在每一个连接处维持着因果链：




不管因果链是有第一环（比如，时间的开端）还是没有第一环（比如，在一年龄无限的宇宙中），上图总是没有问题的。有了上图，我们就可以说，与其说上帝是宇宙的原因，不如说上帝是宇宙的解释。

上述的概念是不容易把握的。粗略地说，物理定律是以事物发生的规律性向我们呈现出来的。行星在其轨道上的精确运行，元素光谱的规则条纹等等都是有规律的。在开动着的汽车上，假如踩了刹车闸，汽车就会慢下来。假如点燃火药，火药就会爆炸。热的火焰会使冰块融化，花瓶落在坚硬的地上会碎裂。这世界并不是杂乱无章无规律可循的，而是至少在某种程度上是可预测的，有秩序的。

我们从受限的时空观点出发，用原因和结果来解释这些规律性，比如，太阳的引力是地球运行轨道发生弯曲的原因，等等。但还有另一种可能性——上帝实际上是所有的事件的原因，上帝从我们的宇宙之外对我们所在的宇宙施加作用，精心地安排事件以显示规律性。

下面是一个有助于说明问题的类比。假设在一机枪手前面有一个靶。该机枪手以恒定的速度扫射枪靶。结果，靶上就出现了间距相同的枪眼组成的图案。这时，某种必须永久生活在枪靶平面上的二维生

物会把这一系列事件看成是在它所在的世界中，孔洞是有规律地出现的。它仔细观察过孔洞之后便会推论：孔洞的形成并不是随机的，而是周期性的，而且，孔洞的布局也有几何学上的质朴性，它们之间的间距是相同的。于是，这二维生物就会自信地宣布一条平面物理的新定律：孔洞创生定律。即每一个孔洞的出现都以一种有规律的方式造成了同行的下一个孔洞的出现。在一简单的序列中，一个孔洞毕竟总是跟随着另一个孔洞出现的。这个二维生物从它那受限的二维世界的观点出发看问题，因而就完全看不到真正的事实：所有的孔洞实际上完全是彼此无关的，其布局的规律性完全是那枪手的行为造成的。同样，宇宙有秩序的运行也可以如此解释：上帝从某种更为广阔的背景以一种有组织的方式创造了每一事件。更为广阔的背景是更高维度的空间？是一种不是空间的物理结构？是一种全然是非物理的结构（这又是什么意思呢）？

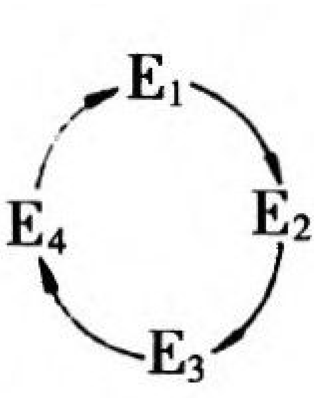
怎样证明这种看法是有道理的呢？看看你的周围吧。看看宇宙的复杂结构和精巧的组织吧。动脑筋想想物理定律的数学公式吧。看看物质的安排，从旋转的星系到熙攘的原子活动，所有的物质都被安排得井井有条，为物质的安排而困惑吧。问一问为什么这些东西会这样吧。为什么会有这样的宇宙，这样的一套定律，这样的物质和能量安排？一句话，为什么竟然有事物存在？

自然宇宙中的所有事物、所有事件，都要依赖于外在于它们的事物才能得到解释。解释一个现象，就得用另一个事物解释它。但是，假如这现象是存在的一切，是整个自然宇宙，那么，按照宇宙的定义，宇宙之外显然没有任何物理性的东西来解释宇宙。因而，宇宙的任何解释就必定是非物理的和超自然的东西。这东西就是上帝。宇宙之所以是这样的，是因为上帝要它这样。科学按定义来说，只是研究自然宇宙的，它可以成功地用一种事物解释另外一种事物，再用其他的事物解释这另外一种事物，如此这般一直解释下去。但自然事物的总体却得从自然之外进行解释。

一切自然事物都要视其他事物的条件而定。以此为基础的推理路线就被称作“偶然性”论证，是上帝存在的宇宙论论证的第二个变体。这种论证如同我们在上一章里讨论过的因果论证一样，也受到了某些批评。

从某种意义上讲，偶然性论证搬起石头砸了自己的脚。我们可以设想把“宇宙”的定义扩大，把上帝也包括进去。那么，又该如何解释由上帝加时空、物质组成的宇宙而构成的整个系统呢？一句话，怎么解释上帝？神学家回答道；“上帝是必然的存在，用不着解释；他自己内部就包含了他自己存在的解释。”但这样的话有任何意义吗？假如有，我们为什么不能用同样的论点来解释宇宙：宇宙是必然的，宇宙在其内部包含了它自己存在的原因？的确，这似乎像是上一章里我们所说的惠勒的观点。

一个自然系统包含着其自身的解释，这在门外汉看来似乎是自相矛盾的，但这思路在物理学里有某种优越性。人们可以姑且认为（不考虑量子效应），一切事件都是偶然的，是要用其他的事件来解释的，但同时也没有必要得出结论，以为这事件的序列或是无限的长，或是终止于上帝。这序列可能闭合成为一个圈。例如，现有四个事件、或事物、或系统，$E_1$，$E_2$，$E_3$，$E_4$，四者可以如下的方式相互依存：



有一种与此一模一样的理论一度在某些粒子物理学家中间流行。这些物理学家想以这样的理论解释物质的结构。下面就是一个著名的解释链：物质是由分子构成的，分子是由原子构成的，原子是由电子和原子核构成的，原子核是由质子和中子构成的。一直有一种很普遍的观点（自古希腊以来就有了），认为这解释链将会有一个终结；有为数不多的真正的基本粒子，浑然一体，不可再分，是一切物质的建筑砌块。假如我们能深入到原子内部更小的区域，迟早会发现这些基本的无结构的粒子。目前，这一理论获得了所谓夸克理论的强有力的实验支持（见第十一章）。

根据不可思议的量子论的特征，则有另一幅图景。量子论认为（这颇为难解，我们将在后面的几章里加以澄清），根本就不存在基本粒子。相反，每一个粒子（至少是每一个亚核粒子）都是由其他粒子构成的。没有哪一个粒子是基本的、原始的，每一个粒子都包含着所有其他粒子的某些个性。若干粒子在一个自洽的解释圈中产生了自己、构成了一个系统，这种想法使人想起了那个掉进泥塘的小孩的故事。那小孩掉进泥塘之后，拽自己的靴襻，硬是把自己拽出了泥塘。因此，物理学家们也把上面的量子论解释法称作“拽靴襻”式的解释。人们可以想象，“拽靴襻宇宙”完全依据自然的、物理性的相互作用而包含了自己的解释。

但是，神学家会反驳说，上帝的能力是无限的，他的知识也是无限的，因而上帝是人们所能想象的最单纯的存在，所以，他比起在很多具体方面都是复杂的和特殊的宇宙来更有可能包含着他自己存在的原因：

很有可能的是，假如存在一个上帝，他就会造出某个有限而复杂的宇宙。宇宙没有什么原因而存在是很不可能的，但上帝却较可能没有什么原因而存在。宇宙的存在是奇怪难解的。假如我们设想它是上帝创造的，就能使它变得可以理解。这种设想比起设想宇宙没有存在的原因，要求解释的开端相对要简单。这就是为什么要相信前一种设

想是正确的。[2]

这反驳是很有说服力的。要想认为有着这么多独特的、可能的特征的精致而复杂的宇宙只是偶然存在着，这是不容易的。我们难道真能把这当作毫无道理可言的、不可解释的事实接受下来吗？于是，一个单一的单纯而无限的精神（尽管其存在在逻辑上就令我们困惑）似乎就是宇宙这个必然的存在物的一个更说得通的解释了。

然而，对无限的精神（上帝）要比宇宙更单纯这样的假设，科学家可能会提出挑战。在我们的经验里，精神只存在于超过某一复杂程度的物理系统中。大脑是个高度复杂的系统（在第六章里我们将要看到，精神必须被当成一个“整体的”概念来看——精神是一种活动的模式）。可以想象一个离体的精神，但必须有某种方式来表现精神活动模式，而精神活动模式本身是复杂的。因而，就可以说，无限的精神是无限复杂的，因此比起宇宙来也就更不可能出现，因为宇宙的很多部件的复杂程度远不足以维持一个精神。

那么，上帝可能不是个精神，而是比精神更单纯的事物了？谈论存在于时间之外的精神有意义吗？思想、决策以及诸如此类的事物不是在时间之中发生的吗？假如上帝不能够作决定（或抱希望、或进行判断、交谈），那么，他对宇宙的性质和存在负责又是什么意思呢？我们是不是可以把不对宇宙负责的存在认作上帝？除了这些疑问之外，我们还得解释宇宙的复杂性和特殊性。为什么会有这样的宇宙？

我将在第十二章里更详细地讨论这个问题，但在这里，我们不妨注意一下，在评价一个包含着自身原因的宇宙和要求以上帝来解释其存在的宇宙这两种观点的相对合理性的过程中，我所认为的中心问题。在前面的讨论中，我们理所当然地认为宇宙是非常复杂的，上帝为宇宙的诸多特征提供了现成的解释。但宇宙从来就是复杂的吗？宇宙的复杂会不会是自然的结果，是由完全普通的物理定律造成的？

根据我们对原初宇宙的最科学的了解，似乎宇宙的确开始于一种最简单的状态——热平衡一一目前所观察到的复杂结构和活动都是后起的。那么可以说，原初的宇宙实际上是我们所能想到的最简单的事物。而且，假如按照表面意义来理解对最初奇点的预测，宇宙就是始于一种无限高温、无限密度和无限能量的状态。至少，这不是与无限精神一样说得通吗？

要想真说得通，关键在于是否能够证明宇宙的复杂性和有序状态真是从简单的原初状态自然发展来的。乍看之下，这种说法似乎是公然与热力学第二定律相矛盾。热力学第二定律的要求与之正相反——有序让位于无序，因而复杂的结构倾向于衰退成无序的最终简单状态。E·W·巴尔尼斯1930年这样写道：

开始时，必定是有最大的能量组织……事实上，上帝一度给钟表（即宇宙机制）上满了弦。假如上帝不再上弦，钟表就会停下来。③

我们现在知道，这样的看法是错误的。宇宙的原初状态不是最大的组织状态，而是一种简单和平衡的状态。这一事实与热力学第二定律的明显冲突只是在最近才获得解决。

原来，热力学第二定律严格地说只适用于孤立的系统。而要想把任何东西从引力那里孤立出来在物理上是不可能的。世上没有引力屏障；即便是有，有关的系统本身也逃不开它自己的引力。在膨胀的宇宙中。宇宙物质都受到宇宙引力场的影响，也就是受到宇宙其他部分的累加引力的影响。有了引力，就开通了通过引力场将有序状态注入宇宙物质的途径。我们知道，假如有外部能量供给，就可以以在一个系统中造成无序为代价，在另一个系统中造出有序。于是，自太阳流入的光和热造成了地球生物圈的高度复杂的有序，但这只能通过不可逆转地牺牲太阳核的有限燃料才能做到。同样，膨胀宇宙也能在宇宙物质中造成有序。

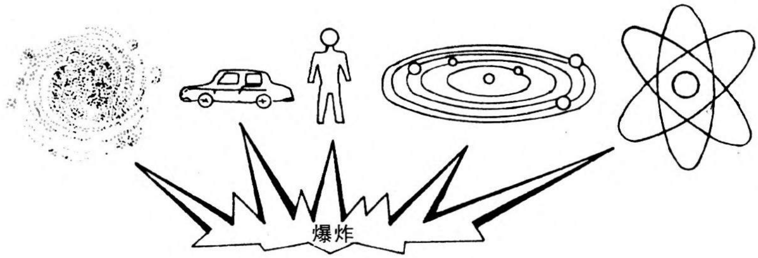



图6　从宇宙的混沌状态中如何生出了有序状态，这是个谜。现今的有序结构和复杂的活动不知是怎样从大爆炸杂乱无章的骚动中产生出来的。显然这违反了热力学第二定律。因为热力学第二定律说，随着时间的流逝，有序程度会降低而不是升高。解决这一难题可能得求助于引力的特异性质。

宇宙的膨胀如何能用来代替上帝“给钟表上弦”，这里可以举一个很简单的例子。我们说过，原初的宇宙物质是非常热的，但宇宙的膨胀使它冷却下来。通过基本的测量论证，可以得出在宇宙膨胀的每一阶段宇宙物质的温度。不过，温度的高低在某种程度上也得看物质本身的特性。辐射热（电磁能量）温度的降低与典型空间区域的膨胀尺度成比例。空间区域扩大一倍，温度就减半。另一方面，诸如氢气之类的物质冷却得快得多。它与尺度的平方成比例。这就是说，只要把氢气跟辐射热分离开来，膨胀宇宙就会在宇宙物质的这两种成分之间产生出温度差。任何一个工程师都知道，温差是有用的能量的理想来源，实质上也是太阳能在地球上产生生命的奥秘所在。因而，宇宙膨胀能够在原先没有有序的情况下产生有序。

利用上面的分析，就可以一步步地追溯到宇宙原初时期的膨胀，从而追究出我们今天在宇宙中所观察到的大部分有序结构的起源。④上面所举的例子实际上不是最重要的。今天的有组织的能量，其绝对最大的来源是高度活泼的氢气，它占了宇宙物质大约75%。氢为所有正常的恒星提供燃料。氢被燃烧之后（在核聚变中燃烧），变成了一些诸如铁之类的较重的元素。铁只是核废料，其中没有可用的核能锁

闭在里面了。因而，恒星有序的存在要倚仗氢多于倚仗铁。

这种情况可用宇宙的膨胀来解释。在原初阶段温度太高，合成的原子核（如铁）不能存在。只有氢核（单个的质子）——最简单的物质——才能存在下去。随着宇宙不断地膨胀、冷却，氢就可以变成一些较重的元素了。正如我们在前一章里讨论过的那样，宇宙物质就是沿着这条路走下去的，但走得不太远。约25%的宇宙物质走到了氦（仅次于氢的最简单的元素），只有极小的部分走过了氦。众多的物质之所以没能走远，这得责怪宇宙膨胀。当初膨胀得太快，使物质来不及经过必要的复杂的核反应变成铁之类的合成重原子核。只有几分钟的“爆炸”，温度就降到了核反应点火所需的温度以下。原子核火堆熄灭了，将大部分物质“冻结”在氢和氦的状态中。只是时间过去很久之后，到了恒星形成时，才形成了分散的热点，使宇宙物质继续向较重的元素进发。

总之，看来在膨胀的宇宙中，有组织的能量可以自然地出现，用不着一开始就有。因而，也就用不着将宇宙的有序状态（低熵）归结为神的活动或归结为在最初的奇点处有有组织的能量输入。奇点可能当初放出来的是完全无规则的、混沌的能量，无规则的能量后来在膨胀的宇宙的影响下将自己自发地组织起来，成为如今的安排。注意，现在我们不仅把物质的起源归结为空间的膨胀（见第三章），而且也把物质的组织的起源也归结为空间膨胀造成的。

不过，到了这一步，事情还没完。最终通过宇宙的膨胀而产生有序的是引力场，而引力场理应因而变得有些向无序方面转化才是。因此，我们虽可以通过把责任推给引力而解释物质性事物何以变得有序，但我们接着还得解释一开始有序是怎么出现在引力场中的；解释到哪里才是头呢？

现在的问题成了热力学第二定律是否既适用于物质，也适用于引力。没有谁真了解这个问题。最近对黑洞的研究显示出某种迹象，热

力学第二定律适用于引力，但不同的物理学家从该研究结果中得出了相反的结论（见第十三章）。有些物理学家如罗杰·彭罗斯得出的结论是，大尺度的宇宙引力场处于低熵（高度有序）的状态，因而这就是说在宇宙创生时得有有序的输入。其他的物理学家如史蒂芬·霍金则认为，宇宙的引力是高度无序的，这是在意料之中的，是受奇点发出的纯粹无规则无结构影响的结果。因为现在还没有人知道怎样将一空间弯曲（即引力）定量化，这一问题悬而未决。然而，两派物理学家的争论表明了一个重要的东西。未来的理论物理学的进展很可能澄清有关的概念，并且能使人们做出有关宇宙创生之时是否有序的确定陈述。但愿科学有一天能回答长久以来为神学家、哲学家所关注的问题。

不管将引力熵如何定量化的争论结果是什么，一个有意思的事已经出现了。在诸如一箱气体这样的系统中，引力是很小的，小到可以忽略不计。低熵（有序）状态是复杂的，而高熵（无序）状态则是简单的。例如，在一个箱子中，假如所有的气体分子都挤在四角上，这显然与平衡（熵值最大的）状态相比是一种较为复杂的状态，因为在平衡状态下，气体是均匀地分布在箱子中的。与此相比，低熵的引力系统从几何学上看要比处于高熵状态的系统简单得多。引力容易自动地生长出结构来。

这样，本来是均匀分布的物质（恒星或气体）会随着时间的过去而结团，形成星团和稠密的积聚物。总而言之，对不受引力作用的系统来说，有序意指复杂，无序意指简单。对引力来说，情况正好相反。

假如宇宙开始时真有一个高度有序的、低熵的引力场，那么，这个引力场就该是平稳的、均匀的。我们看到，在引力这一特例中，同时满足简单的要求和开初的低熵（有序）的要求是可能的。这就是说，我们可以把最简单的宇宙（均匀的宇宙）看成是包含着有无限的可能性，能够产生后来的复杂事物的宇宙。这的确是个令人高兴的结

论。假如我们要相信宇宙的出现是没有原因的，那么，它在开初时物质和引力的形态最好是尽可能地简单，同时又保留了以后演化为复杂和有趣的形态的能力。难道不是这样吗？

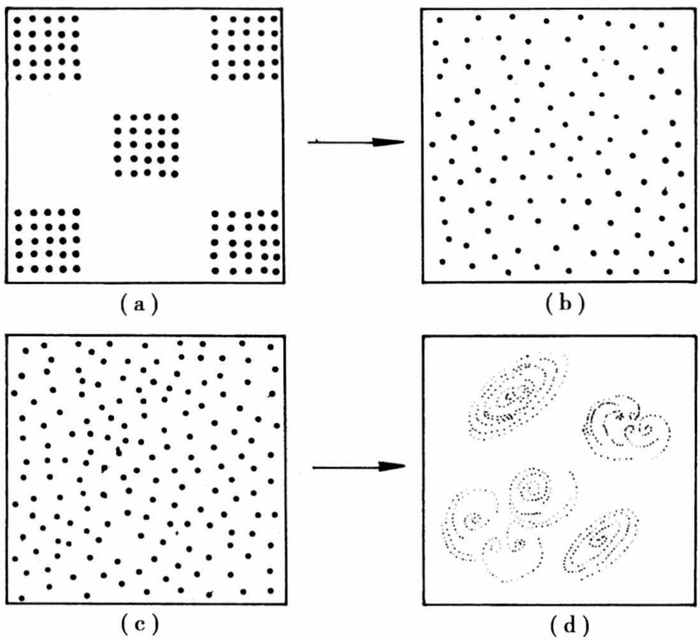


图 7　有序概念的关键是，引力是否可以忽略不计。(a) 箱中盛着一种气体，对这种气体的引力可以忽略。(b) 箱中高度有序的气体很快就由于分子扰动和碰撞而过渡到无结构的无序状态（熵值最大）。其最终状态如（c）所示。与此相比，一种受引力作用的“气体”，如星系，其演变情况却正相反。随着恒星相聚在一起，构成星团（比较星系），起初是均匀的构型（d）就会分裂，结团。结团的最终结果是出现若干黑洞。

前面的那套假说尽管很成功，但问题不光是宇宙的状态，难道定律可以撇在一边吗？至少，即使在开始时宇宙是处在一种非常简单的状态中，物理定律毫无疑问仍是多而特殊的。这些定律难道是因情况而异的吗？我们能够设想一大堆物理定律失效的情况吗？而且，宇宙

的成分——质子、中子、介子、电子等等又怎样解释呢？为什么会只有这些粒子？为什么这些粒子会有它们所具有的质量和电荷？为什么这些亚原子粒子的种类不多不少就这么多？神学家有现成的答案：上帝就要这样。上帝是无限的简单，但愿意创造具有复杂多样性的物理定律和物质的成分，他的目的是创造一个有趣的宇宙。

只是在最近，科学家也开始发觉这类问题的答案。新的进展源自一项理论工作。这项理论工作的目的，是将自然界的诸种力统一在一个单一的描述框架中。根据这一理论框架（我们将在后面的一章里全面地给以描述），现今丰富多样的物理定律纯粹是个低温现象。随着物质温度升高，各种作用于物质的力也开始趋于同一，最后，到了$10^{32}$K这种极高的温度时，自然界所有的力会合成为一种单一的超力，可以用很简单的数学式表示。而且，所有的众多完全不同的亚原子粒子也会失去它们的个性，它们各不相同的特征消失在超高温中。多年来对高能（高能在我们这里就是高温）物理的研究，得出了物质在高温下趋向简单的证据。物理学家们希望发现，随着能量的提高，复杂的亚原子结构会分解，显露出较简单的成分，复杂的力在此过程中也会变得较简单。

假如这些想法是正确的（目前只能说各种迹象令人鼓舞，但要得出什么结论来还为时尚早），那对大爆炸理论会有意味深长的含意。在宇宙创生时的无限高温的情况下，起作用的只有超力和为数不多的几种简单粒子。现今已经分化的各种力和粒子只是随着宇宙冷却才产生出来。这样，宇宙的状态，物理定律，还有物质的各种成分，这一切似乎开初时都处于一种非常简单的状态。

然而，持怀疑态度的神学家会说，即便是可以证明宇宙开初时只有一种单一的超力和为数不多的简单粒子，但这超力和这些粒子又该如何解释？为什么会有这种超力？事实上，为什么会有定律？

这些问题我们要到最后一章再谈。有些物理学家受到自然界基本

定律的简单性的启发，提出一种论点说，很可能最终的定律（在这里是超力）有一种数学结构，这数学结构是唯一在逻辑上没有矛盾的物理原理。这就是说，正像神学家们把上帝说成是必然的一样，在这里，物理也被说成是“必然的”了。那么，我们是不是应当像某些哲学家（如柏拉图）似乎做过的那样，也做出结论说上帝就是物理？

有些物理学家，特别是史蒂芬·霍金，已经提出，实际上我们也该料想到宇宙的原初状态是很简单的。⑤其理由涉及我们在第二章里简略讨论过的原初奇点。奇点的最基本特征是，它像是时空的边缘、边界，因而人们认为它也是自然宇宙的边缘。奇点的一个例子，是一无限致密的状态，标志着大爆炸的开始。据认为，黑洞里也有奇点发生，可能在其他地方也有奇点。

因为到目前为止，我们的一切物理理论都是在时空范围里提出的，所以，时空边界的存在就意味着，自然的物理过程不可能延续到这边界之外。从一基本的意义上讲，根据这种观点，奇点就代表着自然宇宙的外部边界。在奇点处，物质能够进入或离开自然世界，从奇点能发出影响，而这些影响是自然科学所完全不能预测的，甚至原则上也不能预测。奇点是科学所发现的最接近超自然动因的东西。

多少年来，人们曾认为奇点是由于对使用的引力模型过分理想化所造成的一种人工的东西。后来，在一系列精巧和富于启迪的数学定理中，彭罗斯和霍金证明，奇点是相当普遍的；而且，在一切合理的物理条件下，一旦引力变得足够强大，奇点也是不可避免的。在大爆炸中，引力的确足够强大。

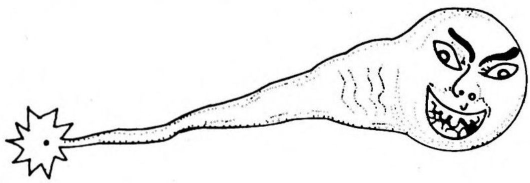


图8　奇点（黑点）代表科学中最终不可知的东西。它是时空的边界或边缘。在奇点处，物质和影响可以以完全不可预测的方式进入或离开自然宇宙。假如奇点是“裸露的”，那么，任何东西都可以在原来没有物理原因作用的情况下，从奇点中产生出来。有些宇宙学家认为，宇宙就是没有原因地从某种类型的裸露的奇点中产生出来的。假如这些观点是正确的，那么，奇点就是自然和超自然的分界面了。

既然人们必须认真地对待他们二人的证明，于是很多人就动脑筋去想奇点会有什么样的行为。各条思路都汇集到了这一点上，即奇点产生的东西或者是完全无序、无结构的，或者是有条理的、有组织的。在前一种情况下，大爆炸的奇点只是产生出一种无秩序的宇宙，没有任何具体的秩序。在后一种情况下，宇宙是带着某种程度的组织出现的，像是钟表上满了弦，可以计时了。

霍金提出了一个“无知原理”，意思是说奇点最终是不可知的，因而也该是完全无信息的（在物理学里，信息大致与有序同义，即负熵）。⑥因此，从奇点出来的任何东西都是完全无序的，混沌的。霍金提出的这一原理，倒是很符合那种认为原初宇宙是处于无序值最大的状态（热平衡）的观点。

很多这类观点，现在都处在现代理论物理学的前沿，因而只能通过理论未来的发展才会得到澄清。物理学家们对时空奇点的问题意见不一，甚至对原初宇宙到底是处于什么样的具体状态也是见解各异。但是，宇宙学最近的进展所产生的各种思想无疑再次引发了关于上帝与宇宙的存在的争论，并且使这一争论转向了一个新方向。

# 第五章　什么是生命？整体论对还原论

“上帝就照着自己的形象造人……”

《圣经·创世记》第1章第27节

“我们是生存机器，是些机器人，被盲目地编了程序，以便保存那些只顾自己的分子，也就是我们所说的基因。”

《只顾自己的基因》理查德·道金斯

神学家认为，生命是最大的奇迹，人类生命是上帝的宇宙总计划中的最高成就。对科学家来说，生命是自然界中最令人惊讶的现象。一百年前，生命系统的起源及演化问题成了科学和宗教有史以来最大冲突的战场。达尔文的进化论动摇了基督教教义的基础。而且，自哥白尼把太阳放在了太阳系的中心以来，进化论又一次使普通百姓感受到了科学分析的深远影响。科学看来好像能够整个地改变人类对自己以及对人与宇宙的关系的看法。

因为本书主要是讲物理学的书，所以我们不会在本书中详述达尔文的革命性理论及他的理论对教会的影响，也不能大讲在最近的“创世论”运动中反达尔文的情绪为何奇异地复活了。这一切问题在其他的书里都有详尽的记述。本章要讲的是科学家对生物体的看法。要讨论的问题是：什么是生命？生命是否为神的精神的存在提供了证据？

圣经相当明确地说，生命是上帝活动的直接结果，生命并不是上帝创造天地之后所确立的正常物理过程的自然结果。相反，是上帝喜欢运用他的能力，先是造出了植物和动物，接着又造了亚当夏娃的。

当然，现在绝大多数的基督徒和犹太教徒都承认创世纪的寓言性质，不想坚持说《圣经》所记载的生命起源就是历史事实。然而，生命的神性，尤其是人类生命的神性，仍然是当今宗教教旨的一个中心问题。

生命是神造的吗？难道上帝真的在奇迹中没有理会物理和化学的定律，把那些由无生命的物质组成的分子捏来捏去，造出了第一个活物？上帝真是在几千年（或几百万年）之前又巧妙地处理了某种猿一样的生物的基因结构，从而造出了人？或者说，生命真是纯粹自然的（或许是复杂的）物理和化学活动的结果，而人是长期曲折的进化过程的最终产品？生命能否用人工的方式在实验室里制造出来？是不是必须得往里面添加一种成分（神的活力），生命才能活起来？

什么是生命？对物理学家来说，生命系统有两个与众不同的特征，一是其复杂，二是其组织。即便是单细胞生物，虽然原始，却显示出任何人工制品所无法比拟的复杂和精巧。例如，我们可以看看一个低级的细菌。仔细观察就会发现，它身上有复杂的功能和形态的网络。这细菌可以用多种多样的方式与其环境相互作用，它可以四处活动，攻击敌人，接近或逃避某些外部刺激，以一种有控制的方式交换物质。其体内的情况像是一个井井有条的大城市。负责进行大部分控制的是细胞核，其中包含着遗传“密码”，就是使这细菌得以进行繁殖的全套化学指令。控制并左右这细菌的一切行为的化学结构是由多达上百万原子组成的分子。这上百万的原子是以复杂的、高度特殊的方式组合起来的。就生命的化学基础而言，最重要的是核酸分子，即大名鼎鼎的“双螺旋”形状的RNA和DNA（核糖核酸和脱氧核糖核酸）。

一个生物学上的生物体是由完全平常的原子组成的。看到这一点是重要的。不错，生物体的部分代谢功能是从其环境中获得新的物质，并排泄变了质的或用不着的物质。活细胞里的一个碳原子、氢原子、氧原子、硫原子，与细胞外的原子没什么两样。这些原子不停地在一切生物体中进进出出。那么，生命显然不能归结为生物体的构成

成分的特性。生命是一种累加现象，像是重量。尽管我们不能怀疑一只猫或一棵天竺葵是活的，然而，假如我们想寻找一种迹象，说明一个单个的“猫原子”或“天竺葵”原子是活的，那我们就得白费力气。

这有时看上去是矛盾的。一些无生命的原子凑到一起怎么就成了有生命的了？有些人认为，不可能从无生命中创造出生命，因此一切活物中必有某种添加的非物质成分，一种生命力。或说是精髓，其本源来自上帝。这就是古老的活力论学说。

经常用来支持活力论的一个论点与行为问题有关。活物的一个显著的特征是，它们的行为看上去都有某种目的性，好像是指向一个具体的目的。这种目的性在较高级的生命形式中最为明显。不过，一个细菌也能表现出它努力达到某些基本的目的，如取食。

在18世纪70年代，卢奇·加尔凡尼发现，用一对金属棒触动青蛙的腿，青蛙腿的肌肉便会抽动。于是，他得出结论说，这种“动物电流”就是那神秘的生命精髓在起作用。的确，那种认为电流与生命力有说不明白的联系的看法，随着弗兰肯斯坦的故事而一代又一代地传下来。弗兰肯斯坦是一个人造怪物，在一台电气装置噼叭响着放电时活了起来。

在最近的这些年里，有一些所谓超自然现象的研究者声称，他们曾以一种相当玄妙的方式把通灵能力与先进技术结合起来，直接探测到了那神秘的生命力。有人出示了一些模糊的照片，上面显示出一些从各种活物身上（其中有人的手指）发出的雾状、丝状射线或光斑。

不幸的是，很难找到任何真正的科学证据来证明这些关于生命力的猜想。显然，那种假想的生命力将其自身显现出来的唯一途径是通过生命；生物显示生命力，非生物则不显示生命力。但这样一来，便把生命力弄成一个词语了，它就不再是关于生命的解释了。我们若说一个人，或一条鱼，或一棵树有生命力，这话是什么意思呢？意思不

过是说这人，这鱼，这树是活的。至于在含糊而神秘的“实验”中所显示的生命力，因这些实验不可重复，所以这些实验及实验结果声名狼藉，显然很容易被指控为欺骗，于是很少有哪个专业科学家把它当真。

乞灵于某种生命力来解释生命的人犯了一个错误。这错误就是忽视了这一事实：由多个部分构成的一个系统可能在整体上会有一些其单个部分所没有的性质，或有一些放在单个部分上便无意义的性质。现在再来举一个例子。报纸上有一张由众多的小点组成的人脸照片。若是只看一个个的小点，看得再仔细也看不出一张人脸。只有把报纸拿开一些，较为粗略地把众多的小点看成是一个整体，才能看出人脸的形象来。这形象不是众多小点本身的一种性质，而是小点集体的性质。人脸是在图形里，而不是在组成图形的一个个小点里。同样，生命的奥秘也不在原子里，而在原子的缔合模式里，其模式即其组合方式。就是说，生命的奥秘在于DNA和RNA之类的分子结构所包含的信息之中。一旦了解了整体现象的存在，也就用不着去寻找什么“生命力”了。原子用不着被“激活”才能产生生命现象。只要将原子以那种合适的复杂方式进行安排，就会出现生命。

我们现在正在讨论的东西有时被称作“整体论”对“还原论”。过去的三个世纪以来，西方科学思想的主要倾向是还原论。的确，“分析”这个词在最广泛的范围中被使用，这种情况也清楚地显明，科学家习惯上是毫无怀疑地把一个问题拿来进行分解，然后再解决它的。但是，有些问题（如钢丝锯）只能通过综合才能解决。它们在性质上是综合的或“整体的”。一幅钢丝锯的图画，就像报纸上由众多小点组成的人脸图像一样，只能在更高一层的结构层面上才能够看得出来，单凭一个个的锯齿是看不出锯来的。这就是说，其整体大于其部分的总合。

在19世纪的物理学中，以及在发展物质的原子理论的过程中，科学的还原论被认真地提了出来。近来，生物学家们沿着还原论的思

路，在阐明生命的分子基础方面取得了许多重大的成就。这些领先的进展鼓励了人们在学问的许多其他领域也采取还原论的思路。

然而，因到处运用还原论而引起的恶果，已招致了尖锐的批评。作家阿瑟·柯斯特勒说，“还原论的看法否定在盲目的力量的相互作用中有价值观、意义和目的的地位，因而这种看法将其阴影投到了科学的界限之外，影响到我们整个文化的甚至政治的气候。”[①]许多批评者说，试图把活的生物体解释成不过是因偶然事件而胡乱结合起来的一堆堆无意义的原子，这严重地贬损了我们的存在价值。

英国神经生物学家唐纳德·麦克埃是位著名的基督教教旨的辩护者。他对在当代生物学家当中盛行的他所谓的“不过是”态度提出了挑战。在《钟表形象》这本书中，为了说明他的论点，他举了一种常见的广告为例。这种广告是由数百只时明时灭的电灯组成的。按顺序时明时灭的电灯组成了讯息。一个电气工程师可以用电路理论准确而完整地描述这电灯组成的系统。然而，若说这广告因而不过是一复杂电路中的电脉冲而已，这说法就是荒谬的。不错，使用电路理论进行的描述，就其本身的描述层面而言，既不错误，也不是不完整；但这描述却没有提到广告所显示的讯息。广告讯息的概念是电气工程师工作范围之外的东西，只有将组成广告的几百只电灯的明灭看成是一整体时才能够显现出来。可以说，广告讯息比电路和电灯高一个结构层面：它是一个整体特征。

就有生命的系统而言，没有谁否认一生物体是由原子集合而成的。错误的观念是把生物体只当作是原子的集合而已。把生物体说成不过是原子的集合而已，这种说法是可笑的，其可笑程度正如把贝多芬的交响乐说成不过是音符的集合而已，或把狄更斯的某本小说说成不过是词语的集合而已。生命的特性、一首乐曲的主题、一部小说的情节，是所谓的显现性的。其性质只有在整体的结构层面上才能显现出来，而在部分层面上则是无意义的。部分的描述与整体的描述并不矛盾，二者是互补的，并且在其自身的层面上是正确的（假如我们仔

细看看量子论，我们就会再次遇到一个系统有两个不同的、互补的描述的概念）。

计算机操作人员很熟悉层面区分的重要性。现代电子计算机是由一复杂的电路和开关的网络构成的，有一系列复杂的电脉冲经过开关。但这只是从硬件层面上进行的描述。另一方面，同一个电路活动可能是代表解出一套数学方程式或分析一枚导弹的弹道。这样的描述就比硬件高一个层面，而且要使用诸如程序、运算、符号、输入、输出这一类的概念，这些概念在硬件层面上是无意义的。计算机内部某一个作为计算机部件的开关是不会通电进行诸如求平方根之类的计算的。它之所以通了电，是因为电压合适，物理定律使然。用程序来对计算机运算进行高层面的描述，这属于软件层面的事。硬件和软件的描述都描述了计算机内部正在进行的活动，每种描述在其本身所属的范围里都是前后一致毫无矛盾的，但二者是处在完全不同的概念层面上。

道格拉斯·霍夫斯塔特写的那本不朽的书《哥德尔，埃舍尔，巴赫》，以最令人信服的方式描绘了还原论与整体论之间的紧张关系。他通过研究一蚁群的兴衰，用他那令人目眩的“蚂蚁赋格曲”清楚明白地揭示出层面混淆所造成的易犯的错误。群蚁具有一种社会结构，既复杂而且有高度的组织。其社会结构是以分工和集体负责为基础的。尽管每一个单个的蚂蚁其行为种类非常有限，或许连某些现代的微处理机都不如，但整个蚁群却显示出很了不起的智能和目的的层面。建造蚁穴牵涉到庞大而复杂的工程。显然，没有哪一个单个的蚂蚁在其脑子里装有这浩大工程的设计图。每一个蚂蚁都是一架自动机，按程序完成一套简单的动作。这就与硬件层面的描述相似。一旦把蚁群看成一个整体，复杂的图形就显现出来了。在这整体层面上（相当于计算中的软件描述），显现的特性，如具有目的性的行为、组织等，是显而易见的。就整体而言，图形显现出来。霍夫斯塔特认为，这两个层面的描述并不冲突。诸如“我们应当通过整体论来理解世界。还是

应当通过还原论来理解世界”，这样的问题被他斥为无用。运用整体论还是运用还原论，这全看你想知道什么。霍夫斯塔特指出，整体论这种看法在东方长期以来就为人们所赏识，并表现在神秘的东方禅宗哲学中。

尽管我们习惯于认为个体的蚂蚁是首要的生物体，但从某种意义上说，整体的蚁群也是一个生物体。其实，我们自己的身体也是一些“群”，是由在集体的组织中上十亿相互合作的单个细胞组成的。众多细胞之间的联系从某种意义上说，要比蚁群中单个蚂蚁之间的联系紧密一些，但就其分工和集体负责而言，其基本原则显然是一样的。然而，这里要注意的关键问题是，正如在一蚁群中存在着显现的整体特性一样，在细胞群中也存在着同样的特性。若说一蚁群不过是一群蚂蚁而已，就是忽视了蚁群行为的实际存在，其荒谬程度就如同说计算机程序不是实在的，因为它们不过是电脉冲而已。同样，若说一个人不过是一堆细胞而已，而这些细胞本身不过是些DNA之类的片断而已，DNA也不过是成串的原子而已，于是就得出结论说生命是无意义的，这些话便是彻头彻尾的胡话。生命是一个整体现象。

对生命的整体特性有了了解，就可以安心地放弃关于生命力的旧观念了，因为这旧观念的由来也是层面混淆所造成的。那种认为必须将某种魔力加到无生命的物质上才能使无生命的物质“活起来”的观念是错误的。其错误就如同认为组成计算机的电开关必须得到“计算力”，计算机才能工作起来，或蚂蚁必须有“群体精神”，蚁群才能有整体功能。假如能够把一个个原子按合适的图形组装起来，人工地造出一个完整的细菌，那么毫无疑问，这人工细菌的各个部分会跟一个“自然的”细菌一样活。

物理学家们早就不再以纯还原论的观点来看待自然界了。尤其是在量子论中，从整体上看待测量行为，对于得出量子论的有意义的解释是很重要的（见第八章）。然而，只是在最近的几年里，整体论哲学才开始对物理学发生较为普遍的影响。这种情况也出现在医学界。

医生们把重点放在治疗“整体的病人”上。心理学家和社会学家也开始接受整体论了。整体论科学于是飞速发展成为某种信仰，可能部分是因它与东方哲学和神秘主义合拍。弗里提约夫·卡普拉的《物理学中的道》以及祖卡夫的《跳舞的物理大师们》很好地描述了这种变化的情况。这两本书都利用了现代物理学和传统的东方整体概念如“太一”之间的类似。

人们一旦承认整体论的观点，从而消除了生命力的必要性，便立刻会遇到一个这样的问题：科学，尤其是物理学，能不能描述包括生命在内的那些整体现象？大卫·玻姆在《整体与暗含的秩序》一书中，试图建立一种范围广阔的物理学。在讨论生物系统时，玻姆说道：“生命本身必须被看成是在某种意义上属于一个整体”。②他继续说，生命是“包含在”一整体的系统中的，这系统包括一些无疑是无生命的部分，如我们所呼吸的空气，空气的分子说不定哪天就被吸收进我们的身体之中。

实际上，一个世纪以前，随着热力学问题的出现，物理学就发展起来以解决整体现象。在此期间，詹姆斯·克拉克·麦克斯韦和路德维希·玻尔兹曼所做的工作就是试图从大群大群分子的统计学性质中推出热力学性质。热力学对生命来说是一门具有头等重要意义的学科，同时也常常使生物学过程看上去具有悖论性质。

这悖论涉及生物的本质，即有序。我们已经说过，调节有序变化的热力学第二定律规定无序总是增加。但生命的发展却是典型的有序增加。在地球的历史上，生命系统演化成为越来越复杂、越来越精致的形式，同样，有序的水平也在提高。这怎么能跟热力学第二定律相符合呢？这是不是一个与第二定律相反的证据，证明神力在起作用，将有序（以奇迹的方式）加在了生物体的发展上？

仔细观察就可以发现，生物学与热力学第二定律之间不见得有什么矛盾。热力学第二定律总是就整个系统而言的。某处有序增加，同

时可能在另一处熵增加。而生命系统的最基本、最重要的特征是，生命系统对其环境是“开放的”。生命系统不是封闭的或自足的。它们只能通过与其环境交换能量和物质才能活下去。计算一下熵的收支情况就可以发现，一个生物体中有序的增长是以更为广阔的环境中的熵为代价的。在所有的情况下，都是熵的净增长。实际上，也有很多无生命系统中有序增加的例子。无形的溶液中析出了晶体，这就说明了一处的有序增加，但仔细研究一下就会看到，在结晶的过程中会补偿性地产生热，而这使得晶体周围的物质的熵增加。

人们普遍认为，生物需要能量，但这种看法不太正确。物理学告诉我们，能量是守恒的，既不能被创造，也不能被毁灭。一个人进食之后，一些能量就在他的体内被释放出来，这些能量然后以热的形式或以活动所完成的功的形式发散到他周围的环境中。一个人身体的总能量大致是不变的。只是有能量流在他身上并流进流出。这能量流是由所消耗的有序能量或负熵所驱动的。那么，用以维持生命的关键要素就是负熵了。伟大的量子物理学家埃尔温·薛定谔在其《生命是什么》一书中这样写道：

生物体具有惊人的本领，能把“有序之流”集中到自己身上，能从合适的环境中“汲取有序”从而使自己免于衰变为混沌的原子。③

认识到生命现象并不违反物理基本定律，当然并不等于说物理定律解释了生命现象。我们只能说物理定律与生命现象不矛盾。没有几个物理学家会说，假如完全了解了原子和分子过程的所有规律，就能够从这些规律中推断出生命会存在。但这并不是要再次犯乞灵于“生命力”的错误：

一位只熟悉热机的工程师，在看过电机的建造过程之后，必会意识到电机是按他所不了解的原理工作的……热机和电机的构造不同，就足以使他意识到二者运行方式是全然不同的。他决不会因看到电机没有锅炉也不用蒸气，一按电门就转起来，就以为电机是幽灵驱动

的。④

同样，有生命的生物体也可能是按目前人们尚未了解的物理定律及过程活动的，尽管人们或许了解组成生物体的各个部件原子和分子。再说一遍，整体行为是不能根据组成该整体的部分来理解的。假如有生命的物质和无生命的物质都遵守同样的物理定律，那么；何以同一套定律会产生出差别如此之大的行为呢？似乎是物质分别进了两个岔道，有生命的物质向着越来越有序的状态演进，而无生命的物质则在热力学第二定律的影响下变得越来越无序。然而，无论是无生命的物质还是有生命的物质，其基本成分原子都是一样的。

近年来，在揭示整体有序状态何以出现的原理方面取得了一些进展。生命的“奇迹”似乎也不那么神秘了。因为人们通过研究发现，一些无生命的系统也能获得自发的组织。这方面有很多例子。可以举一个简单的例子。假如处于水平状态的一层液体被从底下加热，达到一临界温度之后，就会看到该液体自动形成由对流的环形组成的图样，那是由众多的分子按可以看得出来的图样流动形成的。

人们在研究流体的过程中发现，假如使某一系统离开热力学平衡状态，就会产生有序状态。这种例子多的是。流体流动中产生涡流就是这种情况。在地球上，涡流产生了大气环流的模式，导致了龙卷风和其他的大气扰动。在木星上，因为有涡流，木星就有了特色独具的复杂而美丽的表面花纹。

某些化学反应非常清楚地显示了有序状态如何自发地产生。在所谓贝罗索夫一扎布汀斯基反应中，装在试管中的一种化学混合物可以产生水平的条纹，而在一浅盘中则呈现出螺旋花纹。在某些条件下，常可以在有机（但不是有生命的）物质中观察到有组织的化学行为，在很多情况下，其行为涉及高度复杂的包含某种“反馈”和“催化”成分的连锁反应。

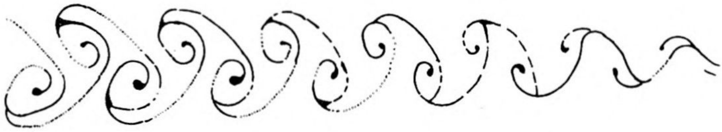


图9　液体流过一细金属丝，产生了如图所示的精巧的涡流，很像是木星表面的花纹（图是根据大卫·特里顿博士的准许复制的）。

诺贝尔奖获得者化学家以利亚·普里高津和他在布鲁塞尔大学庞大的研究组系统地研究了自组织的系统。这里，也应提到曼弗雷德·埃京的开拓性工作。普里高津的目的不光是发现物质自组织的机制，而且他也想用严格的数学方法来描述这些机制。在很多情况下，用于描述高级生物系统中的简单行为模式的方程式与描述无机化学反应的方式相同。普里高津认为，很可能这些流体运动或化学混合物的较为简单的例子，就显示了统辖着生命奥秘的原理。普里高津所举的所有例子都有一个共同的特征，这就是，有关的系统都是远离热力学平衡的。因而，它们都变得不稳定，并且自发地将它们自己大规模地组织起来。普里高津在描述这样的组织时所用的术语是“耗散结构”：“耗散结构出现的一般条件是，有关系统的尺度要超过某一临界值……同时还涉及该系统在其中自始至终作为一个整体活动的长程有序状态”。

毫无疑问，普里高津的研究大大促进了我们对远离平衡的物理结构的理解，也帮助我们认识到，在无生命的系统中也有些很像是生物的模式。然而，拿普里高津的研究结果来穿凿附会却是不智之举。无生命的系统和有生命的系统具有同样的行为，这并不说明二者都可以加以同样的解释。苯环的形状很像是小孩子们玩的牵手游戏，但却不能拿二者之间的相似来解释人类行为。但是，对自组织系统的研究的确表明，生命系统复杂的有序状态，可以很有理由地说是起源于高度非平衡的物理过程。尽管目前人们对生物系统复杂的有序状态所知不多，但也用不着用生命力或神的生气来解释它了。

很多相信宗教的人都准备承认，生命一旦出现在地球上，其后来

的繁衍与发展，就可以用物理和化学定律结合达尔文的进化论给以完满的解释。例如，生物体的生殖，就是DNA螺旋体对自己进行化学复制。这是一种直截了当的机械过程，尽管复杂。但是，生命可是如何起源的呢？

生命的起源一直是个重大的科学之谜。其中心问题是临界值的问题。只有当有机分子变得高度复杂，复杂到一定程度之后，我们才能认为它们是“活的”，意思是说，它们以一种稳定的形式将大量的信息变为密码，不仅显示它们有能力储存复制自己的蓝图，而且它们也拥有实施复制的手段。这里的问题是，普通的物理和化学过程在没有某种超自然作用的情况下，怎么能跨越这一临界值呢？地球的年龄大约是45亿年。化石中所发现的发达的生命迹象至少是35亿年以前的。按理说，在35亿年之前就存在着某种形式的原始生命。用地质学的话来说，当太阳系诞生的阵痛消退之后，生命马上就在我们这刚刚冷却下来的行星上立住了脚。这就是说，不管产生生命的机制是什么，总之这些机制是相当有效的。这促使很多科学家得出结论，认为只要有了合适的物理和化学条件，生命的发生几乎就是不可避免的。

有人提出，促使生命出现的东西是“原初汤”。原始的地球水分充足，水中有简单的有机化合物，是由大气中进行的化学反应形成的。当初的地球，会有无数的池塘湖泊，那里面也会有各式各样的化学反应。经过几百万年，越来越复杂的分子形成了。最后，越过了临界值，复杂的有机分子随机地自组织就产生了生命。

1953年，著名的米勒-尤里实验给这种说法提供了支持。芝加哥大学的斯坦利·米勒和哈罗德·尤里当时试图模拟想象中的原初地球的条件。他们如法炮制了那时的大气——甲烷、氨和氢，还有一滩水，还有闪电（是用放电模拟的）。几天之后，这两位实验者发现，他们的那“滩水”变成了红色，里面有很多对现今的生命来说是重要的化合物，如氨基酸。

这些结果虽令人鼓舞，但并不能因此就认为，这些化合物放在那里，只要经过各种化学反应，过几百万年之后，这些化合物就会自然而然地产生出生命来。只要做一下简单的统计，很快就会发现，有机化合物自发地组成DNA，即带有遗传密码的分子，其几率几乎是意想不到的小。“原初汤”中的分子随机地连结成串组成DNA，这种可能性小得可笑。分子可能的组合方式太多了，偶然碰上一次产生出DNA的机会实际上是零。然而，普里高津的研究表明，如果离开热力平衡，很多系统就能自发地将自己组织起来。因此，原初汤不会老在那里晃荡。某种外来的影响会打破热力平衡，使原初汤进入一系列越来越复杂的自组织反应。这种外来的影响可能就是太阳，其强有力的辐射流使不平衡（负熵）得以产生。驱动着现今地球生物圈的，就是太阳的辐射流。外来的影响也可能是别的东西，没有人知道到底是什么。总之，一系列自组织反应的终端产品可能就是DNA。

总而言之，不难想见原初汤里含有一切生物所必需的成分，由外来的影响驱动着，进入了相互连锁的自组织、自放大的“反馈”循环，由此使有序程度越来越高，极大地增加了越过生命阈的可能性。但是，从米勒-尤里实验过渡到完全能进行复制的分子，这之间有很多中间步骤。若是谁认为我们对这些中间步骤有了了解，那可就错了。生命的起源到现在仍然是一个谜，甚至在科学家中间还引起争论。弗兰西斯·克里克在20世纪50年代揭示了DNA的分子结构，因而被誉为做出了20世纪最重大的发现。即使他，在谈到生命起源时也是谨慎的：

要我们来判断地球上的生命起源到底是一个罕见的事件，还是一个几乎肯定会发生的事件，这是不可能的……那一系列似乎是不可能的事件，要想给其几率一个数值似乎是不可能的。⑤

不过，不了解生命的起源，并不说明生命就是起源于神迹。未来的发现能提供很多我们现在所不知道的细节。

即使将来的研究表明生命的自然起源是个极其罕见的事件，那些

相信宇宙无限、宇宙包含的行星数目无限的人也用不着害怕统计。在一个无限的宇宙中，按纯粹的几率，任何可能的事件都必定在某处发生。显然，我们就是发现我们自己所处的这块区域发生了极其罕见的事件。

对生命、生命的起源以及生命的功能进行的研究是否为上帝的存在提供了证据？我们已说过，现代的科学家把生命看作一种机制，他们找不到关于生命力或非物质特质的任何实在的证据。人们现在仍不了解生命的起源，尽管现在正在兴起的关于自组织系统的研究使某些人觉得有理由认为生命是自动兴起的。生命具有了不起的能力，能够集中负熵，这毕竟并不违反热力学第二定律。当控制和决定生物功能的物理定律仍然有效时，就不会有证据证明有生命的系统实际上与已知的物理和化学定律相矛盾。

这一切当然都没能排除一个造物的上帝存在，但也确实表明，正如土星的光环和木星表面的图案可用物理学来解释其成因一样，生物的发展也不必用上帝的作用来解释。上帝存在的证据要么是到处都有，要么是到处都没有。生命似乎与其他有组织的复杂结构没有什么特别的不同，程度的不同倒可能有一些。我们对生命起源的无知，虽给神造生命的说法留出了广阔的地盘，但主张神造生命实在是一种消极的态度。遇到解释不了的事，就拉上帝来救驾，到头来只能是将来在科学的进步面前退却。因此，我们不要把生命看成是机械的宇宙中的一个孤立的神迹吧。我们该把生命看成是宇宙奇迹的一个组成部分。

科学家们普遍认为，生命是一种物质的自然状态，不过，是一种可能性很小的状态。科学家们的这种看法诱导了人们猜想在宇宙的其他地方或许也有生命。这种猜想当然是一个引起争论的话题，本书不想就它进行评论。直到今天，仍然没有天外生命的肯定性证据，尽管有人说飞临火星的海盗号探测器在其一个实验中确实表明在那里有可能有一种生化反应。不过，仅在我们银河系中，就很可能有几百亿的

行星，因而有些科学家认为宇宙是富于生命的。霍伊尔和克里克就推测，地球上的生命来自宇宙空间。

关于外来生命的可能性的推测，使人想到可能天外有比人类还有智慧的生物。因为地球的年龄还不到宇宙年龄的一半，所以有可能在某些行星上，智能生物几十亿年前就演化出来了。他们的智能与技术同我们相比高得无法想象。具有如此发达能力的生物很可能控制了大片的宇宙，尽管我们察觉不到他们活动的迹象。天外智能生命的存在会对宗教有重大的影响。假如真能找到证据证明其存在，那么，上帝与人有特殊关系的传统说法也就不攻自破了。这对基督教的影响尤其大，因为按基督教的说法，耶稣基督是上帝的化身，他的使命就是拯救地球上的人类。假如有一大群“天外基督”有计划地按时到各个有智能生命居住的行星上去，变成当地的生物形象来实施拯救，这图景未免有些荒唐。然而，如果想不这么荒唐，那天外的智能生命又该怎样拯救呢？

在当今的空间时代，很多人显然都认为确有不明飞行物，但世界上的主要宗教却很少考虑“天外”的事。厄南·麦克穆林是当今少有的几位探讨这一问题的神学家。他认为，“假如宗教在观察上帝与宇宙的关系时，不能够为天外人找到一个位置，那么，在未来的时代里，宗教将会发现越来越难以得到人们的认同。”⑥要是能知道天外的神学家对此问题有什么见解，那倒是挺有意思的事。

在探讨上帝的过程中，我们说到了生命的存在，不管是可以用自然演化加以解释，还是要有神的介入为条件，生命的存在的确提供了有力的证据，证明在宇宙中是有某种目的的。但是，生命只不过是复杂性等级中的一个阶段。生命的重要性在于，它是产生精神的阶梯，也是精神的载体。下面，我们就要讨论精神的问题。

# 第六章　精神与灵魂

“我思，故我在。”

勒内·笛卡儿

“我就是相信，人类自我或曰人类灵魂的某一部分不受制于时间和空间的法则。”

卡尔·吉斯塔夫·荣格

各种宗教尽管对神的本质的看法有很多分歧，我还从不知道有哪种宗教不是说神是一种精神。在基督教里，上帝无所不知，他的知识是无限的。上帝也是无限自由的，想干什么就能干什么。没有什么精神比上帝的精神更博大，因为上帝是至高无上的。

但是，什么是精神呢？

神学家和哲学家长期以来一直在争论这一棘手的问题。然而，对精神的研究今天也进入了科学的领域。心理学，心理分析，最近的大脑研究，计算以及所谓的“人工智能”都在研究精神。这些方面的某些新进展，使我们对自古以来的精神之谜以及精神与物质世界的关系之谜有了全新的看法。这一切，对宗教的影响是深刻的。我们对之具有直接经验的精神是与大脑有关的，计算机是否有精神尚有争论。不过，没有人严肃地主张说，上帝或离世的灵魂也有大脑。完全与物质宇宙分离的精神是很玄的东西，我们暂且不谈，可是，离体的精神是不是有什么意义呢？在本章和下一章里，我们将考察意识、自我以及灵魂的问题。我们要问：身体死亡之后，精神还能不能活下去？

在开始时，我们最好在物质世界和精神世界之间划定一个明确的界线。物质世界充斥着物质的东西，它们有空间的位置，也有诸如广延、质量、电荷之类的特性。这些物质的东西，并不是静止的，而是在四处运动，根据力学规律在变化着、演化着。对力学规律的研究，就构成了物理学的一个分支。物质世界是（至少在很大程度上是）一个公共的世界。所有的人都可以观测它。

充斥精神世界的，则不是物质的东西，而是思想。思想显然不是处在空间之中，而似乎是具有自己的宇宙的，其宇宙是隐秘的，是其他的观察者所观测不到的。思想也可以以各种各样的方式变化、演化、相互作用，也是能动的。对这一切的研究，就构成了心理学的一个分支。

话说到这里，似乎没什么争议。然而，一旦物质世界和精神世界相互作用时，就来了问题。我们的思想宇宙并不是孤立于我们周围的物质宇宙之外的，而是与物质宇宙有着强有力的联系的。我们的精神借助于我们的感官接收到持续不断的信息流。信息流进而产生了精神活动，或是刺激新思想的出现，或是重新组织现有的思想。假如你正在阅读一个句子时，听到屋外呼的一声巨响，你就会不由自主地想，“屋顶上掉下来一片瓦”或“一辆汽车着火了”。因而，物质世界就成了新思想的来源，实际上是重新安排精神世界。

反过来说，精神世界也通过意志现象作用于物质世界。你听到响声后，决定去看个究竟。于是，你的腿动了起来，书也放下了，开了门。你头脑中的思想，借助于你身体这个中介引发了物质的活动，而物质的活动又重新安排了你环境中物质的东西。的确，我们平时在我们的环境中所看到的几乎一切东西，都是精神活动通过物质操作得以实现的结果。房屋、道路、麦田、风车，这一切都始于诸如计划、决策之类的精神活动。它们都是由精神活动变成的“具体实在”。

这一切看起来很明显，但其中已有了一些令人头痛的问题。物质

作用于精神的机制是什么？更令人头痛的是，精神作用于物质的机制是什么？

现在，我们来看看外在的刺激是如何将思想“植入”精神之中的吧。可以那一声巨响为例。声波冲击耳鼓，引起了耳鼓的振动。这振动由三根细巧的骨骼传到耳蜗。耳蜗中的一种膜接收到了振动，再把振动传给内耳中的一种流体。流体又拨动了一些敏感的细丝，受拨动的细丝产生了电脉冲。脉冲沿着神经通道进入大脑。在大脑中，由电脉冲组成的信号碰上了一个复杂的电化学网络，于是，声音就被感知到了。但这是怎么回事呢？这一连串虽说是复杂的物质的相互作用，怎么就突然变成了一个精神事件，变成了对声音的感知？使你真正听到某种声音的大脑电化学模式是怎么一回事？怎么就由听到声音而产生了一系列的思想？

更令人迷惑不解的是，是你听到声音之后的反应。你决定去看个究竟，你的腿动起来。是怎么动起来的呢？大脑的细胞兴奋起来，信息顺着神经传播开去，肌肉紧张起来，你动了起来。

一个物理学家会如何看待你大脑的这种活动呢？首先，他会把大脑的这种活动看成是一复杂电路的作用。各种各样连接着感觉器官和肌肉的神经通道就代表输入和输出的联系。物理学家因为十分了解电路学的规律，所以他会认为，假如能全部了解你大脑的电路（就是说有一份电路总图，另外还可以仔细监测输入信号），那么，经过繁复的计算之后，他就可以精确地预测你大脑这个电网的输出信号，并由此推断出你下一步要干什么。你听到声响之后，会不会出去看个究竟？你大脑发出的电信号会告诉他的。

谁也不会相信，这物理学家真能预测得准。问题在于，假如把大脑看成是一复杂的电路，大脑似乎就完全成了确定性的东西了，因此，至少从原则上看，就成了可预测的了。神经细胞之所以兴奋起来，命令你的腿活动，那是因为大脑电路中的电流模式具有一定的形

式。不同的模式就激发不起神经细胞来，于是，你就会仍旧坐在那里看你的书，不会出去看究竟。

这里难解的问题是，上述那些涉及普通电脉冲的真实的物质事件是与精神事件同时发生的：“这是什么声音？什么东西碎了吗？我该去看看是怎么回事吗？该去。”——大脑细胞就这么开动起来了。尽管到目前为止，对大脑活动的精神描述和物质描述还不矛盾，但其中有一个关键成分却是描述不了的，那就是，你决定去看看那声响究竟是怎么回事。你放下书，你的腿活动，你的一系列活动都是意志的有意识的行为的结果，是选择的结果。在电路确定性的规律中，如何能容得下自由意志呢？

对这个问题的一个解决方法是，可以把精神看作是控制一架复杂机器的操作者。电站的操作者能按下各种按钮，使一座城市大放光明，同样，精神也能使相关的大脑细胞（神经）兴奋起来，按照其决定来驱动身体。但是，要出去看个究竟这种有意识的决定，又是如何使相关的大脑细胞兴奋起来的呢？不是说，电路的规律已经早就决定了输出的信号吗？难道那些规律被违反了吗？难道精神能够进入电子和原子、大脑细胞和神经所属的物质世界，生出电力来吗？难道精神真是能作用于物质，而不把物理学的基本原理放在眼里？难道物质世界的运动真有两种原因，一是平常的物质作用，一是精神作用？

在第十章里，我们将较为充分地讨论那令人迷惑的自由意志和精神物质相互作用的机制问题。但我们的问题还没完。我们仍然没有发现什么是意识、意识是如何发生的。黑猩猩有意识吗？狗有吗？老鼠呢？蜘蛛呢？虫子呢？细菌、计算机有没有意识？人类的8个月大的胎儿有没有意识？1个月的时候有没有？1秒钟的时候呢？对这些问题，怕是没几个人会全回答“有”。那么，意识是逐渐发生的吗？它是一种可以用某种方式定量化的东西吗？我们可以把人的意识定为100，黑猩猩为90，狗为50，老鼠为5，5个月大的胎儿为2，蜘蛛为0.1，等等，这么定行吗？若不然，是不是有一个意识发展临界值，

一过了这个值，意识就猛然剧增，像是某种燃料到了某一临界温度便突然着起来？

我们又能如何识别意识呢？我们每一个人都能直接体验到我们自己的意识。但是，因为我们的意识是处在非物质的思想和感觉的隐秘宇宙中，所以，外人是不能观察到它的。一个人若想推测他人的意识。就只能通过他们的行为和借助于物质宇宙与他们进行交往。张三可以跟李四说，他张三是有意识的，而李四看到张三似乎还算正常，跟自己说起话来还有条理，于是就相信了张三。假如张三是个哑巴，或只能说谁也听不懂的爱斯基摩方言，李四通过观察张三的言行，尤其是通过注意他对刺激的反应、完成复杂任务的表现，等等，仍然会很有信心地得出同样的结论：张三是有意识的。

说到狗的意识，我们可就不能拿得那么准了。狗与人之间的交流是微乎其微的，而且是否真有交流也成问题。狗的大部分行为似乎是无意识的、本能的。可是，没有几个养狗的人会否认他们的爱畜有意识、有头脑，只不过不如（在某种模糊的意义上）人发达罢了。但是，论到较为低等的生物，如蜘蛛之类，就很难说他们有意识了。不错，低等生物也有行为，但其行为显然是自动的，是由本能规定好了程序的。

这样一路思考下来。就容易发现，精神的主动方面和被动方面之间是有所不同的。在有感觉这一意义上的有意识，要比有能力进行计划、做出决定、进行行动这种意义上的有意识低一个档次。一个新生儿无疑能够体验到由身体刺激所造成的感觉，但其感觉几乎完全是被动的。蜘蛛很可能也知道它周围正在发生着什么事，但其反应能力极其有限，只能通过反射运动来做出反应。人们常说，审时度势进而筹划行动是人类独有的行为。这种说法肯定是谬误的（假如有天外智能生命存在，这说法就更错了）。不过，似乎可以这么说：精神的那些较为主动的特性不仅与意识有关，而且与自我意识有关（在下一章里将讨论这个问题）。可能动物的自我概念不大发达。

大型电子计算机的迅速发展，使人们比以往任何时候都更加注意人类思维能力的机制问题。人们对精神与大脑的关系进行了一些仔细的分析研究。研究的中心问题既简单又沉重：机器能够思想吗？

关于所谓的“人工智能”，可说是文献浩瀚，见解纷纭，本书不想对它进行评说。所有的专家至少同意这个观点：现在，即使是最先进的计算机也赶不上人的头脑。大家都知道，在进行算术运算、资料检索、下国际象棋等方面，计算机通常能超过人，但要它们写诗作曲就不行了。这主要是其软件（程序）的问题，而不是计算机的结构硬件问题。大部分计算机都是设计成完成较专门的低级任务的（如进行大量的简单算术运算）。在完成这样的任务时，速度和精确性是最重要的标准。假如一台计算机总出错，不肯出力，还“休假”，或者是行为反复无常，那么，这样的计算机对大多数使用者来说是没什么用处的，尽管计算机有这些荒谬的毛病反倒使它更像是具有人的智能。当然，假使一台计算机有了这样一些人的特点，那么也就没有谁知道该如何为它编制程序了。甚至有没有可能为它编制程序，也没有人知道。同样，人们也不大知道人的大脑是如何工作的。

不管目前的技术局限性如何，机器是否（至少原则上）能有精神这个问题现在仍是个回避不了的问题。用过大型计算机的人很快就会知道，从某种有限的意义上讲，计算机可以以一种类似人的方式与其操作者交往。现代的“交互作用”的技术能够使人与机器进行复杂的问答对话，尽管谈话范围是非常有限的。

我说过，对我们自己的精神之外的其他精神，其存在只能用类比的方式进行推断。假如有人问：“我怎么知道张三有精神呢？”回答只能是：“我有精神，张三的言谈举止跟我一样，也像我一样自称有精神，因此，我的结论是他像我一样有精神。”但这样的推理照样可以用在机器上。你永远也不可能具有他人的精神，也不可能直接体验到他的感觉（假如你能的话，那个他人也就不成为他人了，而成了你了），因而，若设想我们自己之外还有其他精神的存在，这设想就必

然是一种信仰。因此，就“机器能够思想吗”这个问题而言，必定有这样一个回答，即人们不能根据外在的表现（如完成某些需要智能的任务）来断定人比机器高明，因为外在表现只是人可以用来评定机器“内在”体验的外在标准。假如机器被做得像人一样，能够对一切外在影响做出反应，那么，人们就没有可见的理由说机器不能思想或没有感觉了。而且，假如我们愿意承认狗能思想，或蜘蛛、蚂蚁具有某种基本的感觉，那么，即便是现有的计算机在这种有限的意义上讲也可被认为是有感觉了。

1950年，数学家艾伦·图灵在《精神》杂志上发表了一篇题为《计算机与智能》的文章，探讨了机器能否思想的问题。他提出，可以用一个简单的试验来解决这一问题。图灵把这试验称作“模仿游戏”。试验的原理是这样的：让一个男子进入一间屋子，一个女子进入另一间屋子，一个提问者用一种电传打字电报机与这一男一女进行联系，并以问答的方式判定哪一边是男子，哪一边是女子。与此同时，这一男一女都要设法使提问者相信他与她是女子。这样，这男子就必须是一个聪明而娴熟的说谎者。假如用机器来取代这个男子，而机器真能使提问者相信它是个女子，图林就认为这机器是真能思想了。

这种成熟的人工智能是否能实现，对此有很多意见是否定性的。有一派推理认为，计算机是按照严格的理性逻辑方式工作的，因而必然是冷酷的、工于计算的、没心没肺没灵魂又没感情的自动装置。因为计算机的运行纯粹是自动的，所以，它就只能完成作为其操作者的人按程序输入到它里面的东西。没有哪台计算机会离开其操作者，变成一个自主的具有创造性的个体，能够爱、笑、哭叫，具有自由意志。计算机就像汽车一样，是其操纵者的奴隶。

这种推理有一个问题，因为它也能推出个意想不到的结果来。在神经（大脑细胞）层面上，人的大脑也像计算机一样是机械的，也受制于理性逻辑原理。但这并不妨碍我们有举棋不定的感觉，我们照样

会手忙脚乱，兴高采烈，烦闷无聊，不可理喻。

宗教对人工智能这一概念所提出的主要异议是，机器没有灵魂。然而，灵魂的概念却是模糊得要命。早期的关于灵魂的概念是与生命力的概念密不可分的，即灵魂是一种维持生命所必需的、赋予生命的作用。《圣经》，尤其是《旧约圣经》对此论述不多。灵魂的概念似乎主要是从希腊的哲学传统发源的，受的是柏拉图这样的哲学家的影响。《圣经》早先提到灵魂时，是把灵魂当作呼吸和生命的同义词来用的。到了《新约圣经》里，灵魂的概念不知为何鲜明起来，变得与“自我”相当，具有了我们今天所说的精神的那些特点。实际上，“灵魂”这个词在现代用得不多了，现在主要是宗教界使用它。甚至《天主教百科全书》也把灵魂定义为“思想活动的源泉。”[①]因此，灵魂与精神之间的关系是模糊不清的。下面我们将把这两个词当作可以互相替换的词来使用。

宗教教旨的中心观念是，灵魂（精神）是一个事物，肉体与灵魂之间必须划一个清晰的界限。提出这种所谓的精神（或灵魂）二元论的是笛卡儿。笛卡儿的二元论被广泛地吸收进基督教思想中。这种二元论也最符合普通人的信仰。实际上，二元论的观念在我们的文化和语言中是如此根深蒂固，以至吉尔伯特·罗伊尔在其《精神的概念》一书中，把二元论称作“公认的教旨”。

二元论的精神都有哪些特点呢？“公认的教旨”是这样的：人是由两种不同的东西组成的，这就是肉体与灵魂，或说是精神。肉体是精神的寄主或容器，甚或是精神的牢房，精神只能通过上进或死亡才能逃离这牢房。把精神与肉体结合起来的是大脑。精神利用大脑（通过肉体的感官）获得并储存关于世界的信息。精神也利用大脑作为一种手段，通过本章前面所描述的那种方式来作用于世界，实施其意愿。然而，精神（或灵魂）并不驻在大脑或身体的任何其他部位之中，实际上也不驻在空间的任何地方（我在此对某些神秘主义者以及唯灵论者的“非公认的”教旨表示怀疑，因为他们声称亲眼看见过某种与肉体

有着密切空间联系的灵体或灵魂）。

照这样的一幅图景来看，精神的一个重要特点就是：它是一种事物，可能更具体地说就是一种实体。但它不是一种物质实体，而是一种精微的、看不见也摸不着的以太一样的实体，是思想和梦幻的组成材料，是与普通的笨重物质不相干的。笛卡儿对肉体和灵魂的看法，由R·J·赫斯特总结如下：

笛卡儿的基本思想似乎是，首先，存在着两种不同的存在或实在，一为精神，一为物质。精神或精神实在是感官所感觉不到的，在空间中也没有广延性。精神是有智慧、有目的的，其基本特征是思想，或曰意识。②

罗伊尔则说：

虽说人的肉体是一架机器，但它却不是一架普通的机器，因为它的某些活动是由其内部的另一架机器控制的。在内部起控制作用的是一种很特殊的机器，它是看不见也听不见的，没有大小也没有重量的。它不能拆开，一般的工程师也不知道它的运行原理。③

罗伊尔给这内部的机器起名叫“机器之灵”。

灵魂必须具有非物质的特点，似乎有两个理由。首先，我们不能以任何直接的方式看到或观测到灵魂的物质存在，在脑手术中也找不到灵魂。第二，物质世界必须遵从物理定律，而物理定律在宏观层面上（即在忽略量子效应的情况下）是机械的、确定的、因而是与自由意志这种灵魂的基本属性不相容的（这样推理是错误的，我们将在合适的时候说明这一点）。但这些说法只是告诉了我们灵魂不是什么，而不是说明它是什么。于是，我们心里不免犯疑，觉得把灵魂或精神看作是一种事物的看法是没有任何根据的。灵魂之所以给人那么一种虚幻的印象，看上去好像是一种实在似的，其实只不过是因为它顶了几个无意义的名罢了。精神不是机械的，于是乎就是“非机械的”，好

像这形容词对我们有什么意义。罗伊尔的话正是这么说的："精神不是钟表机构，而是非钟表机构。"④

灵魂的属性很难描述，同样，要想找到灵魂在哪里也很难。假如灵魂在空间里找不到，又该上哪里去找呢？（不过，有意思的是，笛卡儿认为，大脑中的松果体是灵魂的所在，或至少是一种结构，为精神和大脑提供了难以捉摸的联系。）新物理学运用它那奇异的空间弯曲以及更高维数的概念，能不能找到灵魂在哪里呢？

我们已经看到，物理学家把时间和空间看成是一块四维的薄片（也可以看作一个气球），他们认为有可能还有其他的薄片，与时空薄片不相连。灵魂是不是驻在另外的某一个宇宙里呢？时空还可以被看作是包在或嵌在一种更高维数的空间里，就像是一个二维的表面或薄片嵌在三维的空间里。灵魂是不是会驻在这更高维数的空间中的某一处，而这更高维数的空间从几何学上讲仍然是与我们所在的自然时空相近，但实际却不在其中？从这更高维数的有利位置出发，灵魂可以在不属于时空的情况下，"附在"时空中的某个人的肉体上。

那些愿意相信离开肉体的灵魂会进入天国的人，还得面临一个更为复杂的安排问题。因为据说，当一个人活在现世时，其灵魂有一种住处，而当灵魂进入天国时，其住处与现世的住处是不一样的。这一类的观念之所以还有人信，就像信几何学的直觉一样，那肯定是因为人们都觉得灵魂该有个住处。所谓灵魂占据着一个位置，就是说它存在于某种空间之中。那空间或是我们平常所察觉到的，或者是我们察觉不到的。假如真是这样，人们就要问一系列的问题：灵魂多大？形状如何？在什么方位？有什么样的运动？而这一切，对灵魂这种由思想组成的非物质的东西是完全对不上号的。

但是，由现代物理学提供的观念还没有用完。我们在第三章说过，有些现代物理学家认为，时空不是原始的概念，而是派生的概念。他们相信时空是由一些次单位组成的（不是位置也不是时刻，而

是一些抽象的实体），这些单位也能体现量子的特点。很可能自然宇宙延伸（在某种比喻的意义上讲）到我们平日所说的时空之外，只有一小部分的那些次单位以一种有秩序的方式结合起来，造成了时空，使时空之外的“其他地方”成了一种互不相连的碎片海洋。这“海洋”难道会是灵魂的属地吗？假如真是这样的话，灵魂就不会占据位置，因为次单位是不会被装配成位置的。因此，对灵魂来说，诸如广延和方位之类的概念是无意义的。实际上，甚至那些诸如里面、外面、之间、相连、不相连之类的拓扑学概念也可能是用不上的。对这个问题，我没有结论性意见。

说到时间，问题就更多了。灵魂不在空间之中，但它在时间之中吗？大概在。假如灵魂是我们感觉的源泉，那么，我们对时空的感觉也必定包括其中。而且，很多可辨的人类精神过程，明显地是依赖于时间的。如计划、希望、悔恨、预期，等等。

假如说灵魂不在时间之中，就会有严重的逻辑问题。假如灵魂超越了前后关系，我们对灵魂在人死后存在还能赋予什么意义呢？在肉体出世之前，灵魂的情况又如何呢？《天主教百科全书》以一种罕见的幽默处理了这个问题：

没有任何证据证明，上帝掌握着大批不属于任何肉体的灵魂，众灵魂只是被上帝注入人类的胚胎之后才归属于肉体……灵魂是上帝创造的，是在被注入物质时创造的。⑤

这里的意思是明显的。灵魂有时候（肉体出生前）不存在。这样的观念显然是与灵魂超越时间的观念相冲突的。

在一切关于永生的讨论中，都贯穿着这同一个基本的时间难题。一方面，人们期望现世生命结束之后，人还能继续存在，不是仅以一种冻结的无时间的存在方式存在，而是还有某种活动。耶稣谈到过“永久的生命”，其内涵就有时间永不结束的意思。

另一方面，这样的概念是与我们对物质世界的时间感觉紧密相连的，因而与所谓的物质和精神分属于不同的世界的说法不大相符。假如有人认为，实际上时间有可能是有终结的（在第十五章将讨论这个问题），很可能根本就没有什么“永恒”，那问题就更麻烦了。

我们在这里所说的这些论点以及其他类似的论点，使很多人感到，灵魂（或精神）和灵魂不死的概念，往好处说是错误的，往坏处说是前后矛盾的。

哲学家们探讨了好几种二元论的替代品。有一个极端是唯物论，完全否认精神的存在。唯物论者认为，精神的状态和作用不是别的，正是物质的状态和作用。在心理学领域里，唯物主义变成所谓的行为主义，宣称所有的人都以纯机械的方式对外界的刺激做出反应。另一个极端就是唯心主义哲学，认为物质世界并不存在，一切都是感觉。

我觉得，二元论的错误在于试图用一种实体（精神）来解释一种实际上的抽象概念，而不是解释一种客体。在科学史上和哲学史上，人们一直喜欢情不自禁地把抽象概念还原成事物，如燃素、热的流体理论、传光的以太以及生命力。这都是些已被证明为不可信的概念。与这些概念相关的现象，需要用抽象的东西，如能量或场来解释。

一个概念是抽象的而不是实体的，这并不会使这一概念变得不实或虚幻。一个人的国籍既不能称也没法量，也不在他身体内占据一个位置，然而，国籍是有意义的，是这个人的重要组成部分。那些不幸没有国籍的人太明白这一点了。效用、组织、熵、信息这类的概念也不是物体意义上的“东西”，而是物体之间的关系，物体的条件。

二元论的基本错误是把肉体和精神看作是一个钱币的正反面，而二者实际上是属于完全不同的范畴的。精神和精神与肉体关系问题上的一切混乱和矛盾，在罗伊尔看来都是这种范畴错置造成的：

按照一种逻辑说，存在着精神；再按照一种逻辑说，存在着肉体；这两种说法都完全正确。但是，这两种说法并不是指有两种不同的存在。⑥

说“存在着岩石”跟说“存在着星期三”都是正确的，但是，若把岩石跟星期三并列起来，并讨论它们之间的相互关系，那就没有意义了。我们也可以用一个罗伊尔的类比，这就是，探讨英国下议院与英国宪法之间是否有过会话是荒谬的。因为二者分属于不同的概念层次。

罗伊尔的这番话，抢先说出了很多近年来才由“整体论”探索的东西。我们在上一章里讲过，精神与肉体的关系，类似于蚁群与单个的蚂蚁之间的关系，或一部小说的情节与字母表的26个字母的关系。精神与肉体并不是一个二元性事物的两个部分，而是分级描述中的两个不同层面的两个完全不同的概念。我们又一次碰上了整体论对还原论。

只要人们认识到，在没有任何神秘的外加实体或成分的情况下，抽象的、高层面的概念照样可以与作为它们基础的低层面的结构一样真实，那么，很多旧有的二元论难题也就迎刃而解了。物质用不着加进生命力才能变得有生命，同样，物质也用不着有灵魂实体，才能变得有意识：

我们这个世界有很多东西，既不神秘莫测，也不是单纯地由物理砌块构成的。你相信人的声音吗？理发又如何呢？有这样的事吗？它们都是什么呢？用物理学家的话来说，什么是一个孔洞呢？这里不是说那不可思议的黑洞，而只是一块奶酪上的洞。这洞是自然的东西吗？什么是交响乐呢？“星条旗”这首歌在时空中的什么地方呢？这首歌是不是就是国会图书馆里的哪张纸上的一些墨迹？把那张纸毁了，美国国歌依然存在。拉丁语依然存在，但已不是活的语言了。法国洞穴人的语言已经不存在了。桥牌这种游戏才存在了不到100年。它是

种什么东西呢？不是动物，不是植物，也不是矿物。

这些东西都不是具有质量的物体，也不是一种化合物；但它们也不是纯粹抽象的东西，就像无理数π一样，是可变的，在空间和时间中是找不到它的位置的。这些东西也有出生地和历史。它们可以变化，可以发生什么事。它们可以四处运动，就像是一个物种，一种疾病，或一种流行病。我们万不可以认为，人们想认真对待的任何东西都可以被认为是在空间和时间中四处活动的粒子群。这绝不是科学的观点。有人或许会认为，你不过是一特定的活的物质的生物体，是一堆活动的原子。这些人或许以为这是常识，或是很好的科学思想。其实，这种想法正说明，这些人缺乏科学的想象，倒不缺乏精明的牵强附会。自我有一种同一性，超越任何具体的活的肉体。人们不必先得相信有灵，才能相信有自我。⑦

大脑是由几十亿神经元组成的。这些神经元整天忙来忙去，也不知道有个什么总体规划（就像上一章里所讨论的蚁群中的群蚁一样）。这就是大脑电化学硬件的物质的、机械的世界。另一方面，我们有思想，有感觉，有感情、意愿，等等。这高层面的整体的精神世界也同样不知道大脑细胞，我们在完全不知道神经元协助的情况下，可以舒舒服服地进行思想。但是，大脑的低层面是由逻辑统治着，高层的精神面却可以是非逻辑的、有感情的，这二者并不一定矛盾。关于这种神经—精神的互补，霍夫斯塔特曾给以生动的说明：

比如说，你现在正犹豫不决，不知道该要一份奶酪三明治还是要一份菠萝三明治。这是不是说你的神经元也在犹豫不决，不知道要什么好？当然不是。你打不定主意要什么样的三明治，这是一种高层面的状态，而这种状态完全是以成千上万的神经元按非常有条理的方式有效地工作为先决条件的。⑧

用一个类比就是，一部说得过去的小说是由一系列语法构造组成的，而这些语法构造又是符合语言以及表现方式中相当严格的逻辑规

则的。然而，这并不妨碍小说中的人物恋爱或大笑，或以一种完全不守规矩的方式行动。假如说这本书是由逻辑的字词构造而成的，所以这故事必得是严格符合逻辑规则的，这样说就是荒谬的，因为这是把两个不同的层面搞混了。麦克埃在探讨神经活动与精神活动的时候，也强调了要避免层面混淆的重要性：“同一个情况或许得有两种或两种以上的解释，而且每一种解释在自己的逻辑层面上都是完整的。这话听起来抽象难解，但正如我们已经看到的那样，其实是可以用无数的例子来加以说明的。”麦克埃在讨论描述层面的问题时，举了由闪亮的灯光组成的广告为例。广告的图像文字完全可以用电路理论来解释描述，但麦克埃指出，还有一个与之互补的描述，即商业信息：“假如给以正确的区分，两者（两种描述）就不会矛盾，反而互补，就是说，每一种描述都揭示了应当重视的一面，而这一面是在另一种描述中没有提到的。”

精神也与此类似：

戴拉·德夏尔丹之流的作家宣扬说，假如人是有意识的，那么，在原子中就必定有一些意识的踪迹。这种说法是相当缺乏理性的基础的……意识与物质粒子不同。对物质粒子的行为，我们可以争论出个不承认也不行的结果。但对意识，我们就不能指望有这样的争论结果。⑨

用更为现代的用语来说，精神是“整体性的”。

上述的一切，当然都不排除人造精神、思想机器等等的可能性。奇怪的是，很多人乐于承认他们的爱畜有智慧，但一想到计算机会有智慧就心惊肉跳。或许，这是一种自我中心的反应，觉得哪一天计算机有了比我们的智力还强的精神，是对我们的威胁。或许，这是一个更为微妙的问题。

对精神和肉体进行两个层面（或多个层面）的描述，这比起原先

的二元论（精神和肉体分属两种不同的实在）和唯物论（精神不存在）是很大的进步。多层面描述如今是一门哲学。随着认知科学，如人工智能、计算科学、语言学、控制论、心理学等等的出现，多层面描述也迅速流行开来。上述的学科都涉及信息处理系统。信息处理系统可能是人，也可能是机器，可以用这种方式处理信息，也可以用那种方式处理信息。与计算机有关的语言和概念的提出，如硬件和软件的区分之类，为了解思想和意识的本质开辟了新的途径。于是，科学家们在思考精神问题时，不得不比以往更加不含糊。

与这些科学进展同样重要的是，一种关于精神的新哲学出现了。这种哲学与上面所说的那些观点密切相关，被人们称作机能主义。机能主义者认识到，精神的本质要素不是硬件（不是构成大脑的物质，也不是大脑所利用的物质作用），而是软件，是大脑物质的组织，或说是其“程序”。机能主义者并不否认，大脑是一台机器，神经元纯粹是按照电路原理工作的，大脑中的物质作用并没有精神的原因。但他们仍是认为，大脑的各种精神状态之间是有因果关系的。粗略地说就是，思想引发思想，尽管在硬件层面上。因果关系早就固定了。

大多数计算机程序编制人员理所当然地认为，硬件层面和软件层面的因果关系没有可比性。他们会异口同声地说：“计算机是由一些电路组成的，它能干什么事是由电学定律规定的。它的输出，是它走过事先规定好的电路的自然结果”。说完这样的话后，他们就会谈论计算机解方程，作比较，作决策，根据信息过程得出结论，例如进行推理。这样，我们就可以有两种不同层面的因果关系，即硬件和软件，而与此同时，我们也不必追究软件如何作用于硬件。精神如何作用于肉体这个古老的话题，于是便可以看作是概念层面混淆造成的。我们从不会问这样的问题：“计算机程序如何使计算机电路解方程”？同样，我们也不必问思想如何开动神经元去产生肉体反应。

机能主义对宗教意味着什么呢？

机能主义似乎类似双刃剑。一方面，机能主义否认精神是人类独有的，声称机器也能思想，有感觉，至少原则上如此。而传统的观念则认为，上帝将灵魂赋予人，于是机能主义与传统的宗教观念难以相容。另一方面，机能主义把精神从人的肉体这一牢笼中解放出来，从而使永生的问题成为悬而不决的了：

对精神进行软件层面的描述，从逻辑上说不需要神经元……这种描述为离体的精神存在留下了余地……机能主义并不排除有可能（不管这可能性多么微小）存在一种机械的灵气系统，这种系统具有精神状态和作用。⑩

机能主义一举解决了关于灵魂的大部分传统问题。灵魂是由什么材料构成的？这样的问题是无意义的，其无意义就像是问公民权或星期三是由什么材料构成的。灵魂是一整体概念，根本不是由什么材料构成的。

灵魂在哪里呢？哪里也不在。以为灵魂在一个什么位置，就像是想找出数字7在什么位置，或贝多芬交响乐在什么位置一样，都是理解错误。像灵魂、数字这样的概念在空间里是找不到的。

时间与灵魂的问题又如何呢？某种东西，只在时间中存在，却不在空间中存在，有这样的事吗？

这个问题更难捉摸。我们常说失业率升高，时尚发生变化，意思是说这些事物依存于时间，却不能固定在某一位置上，不因此而有意义。同样，也完全可以说，精神依存于时间，却在空间里找不到。

因而，我们可以说，那种认为精神不过是大脑细胞的活动而已的看法是错误的，因为持这种看法就是陷入了还原论的泥坑。不过，精神的存在似乎的确是仰仗着大脑细胞的活动的。于是，问题就来了：离体的精神能够存在吗？可以再打一个类比，一部小说是由字词组成

的，但小说也完全可以以声音的形式储存到磁带上，或译为密码用穿孔机打在卡片上，或变成数字存到计算机里。大脑死亡之后，精神能够转移到另一个机制或系统里，继续活下去吗？显然，这在原则上是可能的。

然而，大多数人并不为他们死后整个个性是否还继续存活而操心。因为我们的生活在很大程度上是与我们肉体需要和性能联系在一起的。例如，性活动假如没有肉体或生殖的需要就是荒唐可笑的。很多人也不想要他们的个性中的消极面，如贪欲、嫉妒、仇恨，等等，不想要这些消极的东西在他们死后继续存活。于是，他们希望精神的内核若是在人死后继续存活，就必须去掉其明显的肉体因素以及那些令人不快的特征。但这样一来，灵魂还剩下些什么呢？人的自我又会如何呢？

# 第七章　自我

“每一个自我，都是神的创造。”

约翰·艾克利斯爵士

“我一生中有一件憾事，这就是我不是另外一个人。”

伍迪·艾伦

我们是谁？我们每一个人在我们意识的深处，都埋藏着一种强烈的自我意识。我们成长、发展，我们的思想和趣味发生变化，我们对世界的看法变动不居，新的感情时有出现。然而，在这一切变化的同时，我们从未怀疑我们还是同一个人。我们体验了这些变化。这些体验发生在我们身上。但是，具有这些体验的“我们”又是什么呢？这，就是长久未解的自我之谜。

跟别人交往时，我们通常把他们与他们的身体等同起来，而且在较小的程度上也与他们的个性等同起来。但我们却以相当不同的方式看待我们自己。当一个人说“我的身体”时，他是把他的身体当作一种所有物，就像是说“我的住所”一样。说到精神时，情况就不同了。精神不是一个人的所有物，而是一个人的所有者。我的精神不是一种所有财产，而是我本身。

于是，精神就被看作是体验和感情的所有者，是思想的中心或焦点。我的思想和我的体验属于我自己，你的思想、体验属于你。用苏格兰哲学家托马斯·里德的话说就是：

不管这自我是什么，反正它能思想，会思考，能做出决定，能行动，也能感受痛苦。我不是思想，我不是行动，我不是感觉；我是某种能思想、行动、受难的东西。①

神学家们把自我等同于难以捉摸的精神实在或曰灵魂，这难道不是再自然不过的事吗？况且，灵魂不在空间里，因而不能拆开、散播，于是，自我的完整性也就有了保证。可以察觉到的“自我”的最基本的性质之一就是，它是不可分割的，是分立的。我是一个个人，我与你界限相当分明。

精神（或灵魂）的概念，正如我们在前一章里所看到的那样，是出名地晦涩难解，有时还自相矛盾。“我是什么”？这个问题不容易回答。正如罗伊尔所指出的那样，“我们一旦伸出头去，查找以代词‘我们’命名的那些人时，那无缘无故的使人迷惑的事就来了。’②然而，要想理解永生，就必须解决这个问题。假如我死后继续存在，那么，我能指望那继续存在的是什么呢？

据大卫·休谟说，自我不过是一堆感觉而已：当我最接近我所谓的自我时，我总是碰上这种或那种感觉，如冷与热，明与暗，爱与恨，痛苦与快乐。我在任何时候都会发现“自我”有某种感觉，而且，除了感觉之外，我也不会观察到任何东西。③

假如照这种哲学观点看问题，那么，如果问“我是什么？”回答就是“我是我的思想和感觉。”然而，这样的回答让人不安。难道没有思想者，思想能存在吗？又如何将你的思想同我的思想区别开来呢？而且，所谓“我的思想”究竟是什么意思呢？实际上，休谟后来就其自我不过是一堆感觉的观点，又这样写道：“更严格地省察关于自我的那一部分之后，我发现自己进了迷宫。”

不过，必须承认，自我的概念是模糊不清的，而感觉对形成自我的特质也是很有作用的，即便感觉不能完全解释自我。自我的某些方

面似乎是处在自身同一性的边缘。例如，我们要到什么地方去找（比喻意义上的“找”）情绪？换言之，你的情绪难道不是你整体的一部分吗？众所周知，情绪是很受物质的影响的，例如，血液的化学成分就影响情绪。激素不平衡能导致各种各样的情绪紊乱。药物能够导致也能够抑制多种精神状态和情绪，任何饮酒的人都知道这一点。更厉害的是，大脑手术能够导致一个人个性的重大变化。这一切，都使我们不乐意让灵魂带上太多个性的累赘。另一方面，假如一切情绪都被消除，那还会剩下什么呢？基督徒或许赞成甩掉消极的情绪，但希望灵魂保留爱与敬神的感情。道德上中性的感觉，如厌烦、兴致勃勃、幽默感之类，其去留则有待争论。

人们更为关注的问题是记忆以及整个的时间感觉的问题。我们对我们的自我的看法，是与我们对过去的记忆分不开的。人们还不清楚假如没有记忆，自我能否有任何意义。或许有人会说，一个患健忘症的人虽然可能对“我是谁”产生疑问，但他却在任何时候都不怀疑有一个“我”，而那个“谁”是属于我的。不过，这只是证明，健忘症患者并没有完全失去记忆。比如，在使用日用品方面，如使用杯、盘、公共汽车、床时，健忘症患者并没有困难。而且，他的短期记忆仍是完好的：假如他决定去花园里散步，他是不会一到花园没一会儿，就不知道他到花园是干什么来了。

假如一个人真是完全丧失了记忆力，连几秒钟以前的事也记不住，想不起来，那么，他的自我意识就会完全崩溃。他甚至连动作连贯也做不到。他的身体动作不会有任何意识来协调。他完全不能理解他的感觉，也不能将他对周围世界的感受理出头绪。他自己这一整个的概念与他所感觉到的世界之间的区别，也将会混乱。他将看不出各个事件的规律性或模式，连续性的概念，尤其是他自身的连续性概念也不能维持了。

因而，我们之所以能感到自身同一性，之所以每天能认出我们自己不是别人，在很大程度上是因为我们有记忆。我们在一生中。都驻

在同一个肉体之中，但肉体可能发生相当的变化。肉体中的原子因新陈代谢活动而有条不紊地被取代，肉体成长、成熟、衰老、最后死亡。我们的个性也会有大变化，然而，尽管我们在不停地变形，我们仍然相信我们依然是同一个人。假如我们不记得我们先前的状况，那么，“同一个人”这个概念除了表示肉体的连续性之外，还会有什么意义呢？

我们现在不妨设想，有一个人自称是拿破仑再世。假如他长得不像拿破仑，那么，可以用来判定他的说法真伪的唯一标准就该是他的记忆。拿破仑最喜欢的颜色是什么？在滑铁卢大战之前，他有什么感觉？他该讲出一些关于拿破仑的具体的（并且是完全可以证实的）事来，然后，你才能相信他真是拿破仑再世。然而，假如这人说他已完全丧失了关于他的前世的记忆，只记得他是拿破仑，那你该怎么办？他说“我是拿破仑”这话有什么意义呢？

他很可能会跟你争辩说，“我的意思是，尽管我的肉体、我的记忆、还有我全部的个性现在都属于约翰·史密斯，然而，约翰·史密斯的灵魂的确就是已故的拿破仑·波拿巴的。我以前是拿破仑，现在是史密斯。但拿破仑和史密斯都是同一个我。只是我的特征改变了。”他这话难道不是胡言乱语吗？要想把一个人的精神与另一个人区别开来，不就是要看他的个性和记忆吗？若说存在着某种可能转移的标记，即灵魂，而灵魂除了表现为一种神秘的记号之外又没有什么特性，那么，这样的活便是毫无意义的猜测。要是有人否认有这样一种标记，那该怎么反驳他呢？假如真有这样的一种标记，我们岂不可以说万物都有灵魂了吗？植物、云彩、岩石、飞机岂不都有灵魂吗？有人很可能说，“这机车看上去是一台普通的内燃机车，但它实际上带着的是史蒂文森的那台火箭号机车的灵魂！尽管二者式样不同，材料不同，这内燃机运行时一点也不像那火箭号机车，但它确实就是那台火箭号，只不过结构、外表、设计完全不同罢了。”这样的空洞的话有什么用处呢？

我们可以再举一个比拿破仑再世更平实的例子。假设一位亲密好友接受了一场大手术，手术后面容体貌完全改观，认不出来了。你怎么知道他与手术前的那个人是同一个人呢？假如他告诉你他先前的事，让你想起以前的一些小事，还有以前你与他之间的私下谈话，而且，他还表现出他对以前的情况了如指掌，那么，你就会得出结论说，他的确就是你以前的那位亲密好友。“没错，就是他，别人不可能知道那些事。”但是，假如那场大手术也使你的这位朋友失去了很多记忆，或干脆损坏了他的记忆，那么，你就很不容易判断他是谁了，假如他一点记忆也没有了，你就没有任何根据说眼前这个人是你的朋友（有点根据的话也是他身上还残留了一些原先的体貌特征）。实际上，一个人假如没有任何记忆能不能算是个人还不清楚，因为他一点具有连续性的特征也没有。假如他没有如个性之类的连续性特征，我们通常就不能把他看成是一个“个人”。没有记忆的人对外界的反应或者是完全没有规律的，或者是纯粹的刺激反应，因而，他的行为跟一个程序没编好的自动装置没什么两样。

在这里我们看到，相信人死后灵魂继续存在的二元论者面临着明显的困难。假如灵魂得靠大脑来储存记忆，那么，肉体死后，大脑怎么还能记事呢？假如肉体死后，灵魂什么也记不住了，我们又怎么能说它跟某一个人是同一的？是不是可以认为，灵魂有某种非物质的替代记忆系统，这记忆系统与大脑平行地发挥作用，但离开了大脑也照样能工作？

有人为了试图摆脱这种难题，主张灵魂是超越时间的。灵魂不在空间之中，同样，它也不在时间之中。但这么一来，又引起了一大串难题，我们在前一章里已经看到了。

假如我们注意一下很多哲学家的一个观点，我们似乎就能接近于理解自我。哲学家们的观点是：人类的意识并非仅仅是由知觉构成的，而是由自觉构成的。我们知道我们知道。1690年，约翰·洛克就强调指出，“谁想感知什么事，而对他自己有感知这一点却没有感

知，这是不可能的。”④牛津的哲学家J·R·卢卡斯也持有相同的观点：

说一个有意识的人知道某事，这不仅是说他知道该事，也是说他知道他知道，而且他知道他知道他知道……意识之所以会出现悖论，这是因为一个有意识的存在能够知道自己，也知道其他的事物，然而它却不能被解释为可分的。⑤

同样，A·J·艾耶尔也写道：“人们很想把人的自我看成是中国的套盒，每一个盒都俯视着套在它里面的那个盒。”⑥

要想解开精神之谜，自指（self-reference）的性质无疑是一个关键。在普里高津的耗散结构里，我们已经看到了反馈和自联结（self-coupling）的重要性。耗散结构具有自组织的能力，而自然界中似乎存在着从无生命到有生命到有意识这样一个复杂性和自组织的等级排列。但在这等级排列之中还隐藏了另一个等级排列，这就是上一章里所讨论过的概念层面之分。生命是一整体概念，而还原论者的观点只揭示了我们身体中的无生命的原子。同样，精神也是一整体概念，是属于另一个描述层面的东西。我们不能通过大脑细胞来了解精神，正如我们不能通过构成细胞的原子来了解细胞。在大脑细胞之间寻找智能或意识是徒劳的。在脑细胞的层面上，意识的概念是无意义的。那么，自知的属性显然是整体性的，在大脑具体的电化学机制中是找不出来的。

对自指的研究总是碰上悖论，不仅是在自知这一哲学问题上，而且也在艺术上，甚至在逻辑和数学层面上都碰上了悖论。古希腊学者埃庇米尼底斯使人们注意到自指陈述的问题。通常，我们认为每一个有意义的陈述或是为真，或是为伪，二者必居其一。但我们且来看看埃米尼底斯的命题（我们称之为A），其命题可翻译如下：

A：本陈述是伪的。

A是真是伪呢？假如是真，可陈述本身却说它是伪的；假如是

伪，那陈述肯定就是真的。但A不可能同时又真又伪。因此，A是真还是伪这个问题是没有答案的。

我们在第三章里所见到的罗素的悖论，也与此类似。二者都是由说得通的陈述或概念组成的悖论。只是这些陈述或概念套成环，指向它们自己，就出现了荒谬。A还有一个形式：

A：下面的陈述是真的。A1

上面的陈述是伪的。A2

在这个形式中，A1和A2这两个陈述单独看都是完全平铺直叙，毫无悖论的。但二者一连结起来，组成一个自指的环，再看上去就成了没有逻辑意义的东西了。

在他那本著名的书中，霍夫斯塔特指出了荷兰艺术家M·C·埃舍尔如何在其作品中戏剧性地表现了“局部”有意义的概念圈成了“全局的”悖论。例如，在《瀑布》这幅画中，假如我们顺着圈中的水流看下去，水流在每段路上似乎都是完全正常的、自然的，但最后，我们突然惊讶地发现，我们又回到了开始的地方。作为整体来看，整个的圈显然是不可能的，但圈上的每一段却都“没问题”。呈现悖论的是全局的或整体的方面。霍夫斯塔特在巴赫的赋格曲中也发现了相应的音乐“怪圈”。

一些关注数学逻辑基础的数学家和哲学家，对自指的问题进行了深入的探讨。其中最惊人的成就很可能要算德国数学家库尔特·哥德尔所证明的一个结果，叫作“不完备定理”。霍夫斯塔特的书的贯穿性主题，就是这个定理。哥德尔的定理，起因于数学家们试图将推理过程系统化，以便为建立数学大厦清理好基础。例如，罗素的悖论就起因于试图以一种尽量普遍和尽量不明确的方式使众多的概念分属不同的“集”，以便使概念组织起来。他的这一努力导致了灾难。

哥德尔偶然想到了用数学符号来编排陈述。这种做法本身并无新奇可言。任何一个看过枚举简算的人都会做。哥德尔所探索的新奇的东西是用数学来编排关于数学的陈述，这就又出现了自指。于是，很可能是必然的，类似于埃庇米尼底斯悖论的东西出现了，不过这次是一条关于数学的陈述，实际上是关于那老掉牙的数字1，2，3，……的陈述的。哥德尔在其定理中证明，总是有一些关于数字的陈述，这些陈述在一有限的公理集的基础上，永远也不能被证明为真或为伪。甚至在原则上也不能证明（就像上面所说的A一样）。所谓公理，就是你不加证明就假定为真的东西，如1＝1。这就是说，即使像数论这样相对简单的一个数学系统也具有一些性质，这些性质不可能在一有限的公理集基础上被证明（或被否证），不管那些公理有多么多，多么复杂！

哥德尔的不完备定理的重要性在于，它通过把主体与客体混在一起，证明了在逻辑分析的基本层面上，自指能够导致悖论或不决。这定理现在也被认为是意味着，一个人永远也不能了解他自己的精神，甚至原则上也不能了解。霍夫斯塔特推测道；“哥德尔的不完备定理很有那个古代童话的味道。那童话警告人们说，寻求自我认识就是踏上……永远也完结不了的旅程。”⑦

哥德尔的定理也被人们用来说明精神的非机械性质。在一篇题为“精神、机器和哥德尔”的文章中，卢卡斯断言人的智能是计算机永远也达不到的：“在我看来，哥德尔的定理证明了机械论是错误的，即精神是不能像机器一样解释的。”他的这一论点的核心是，我们人能够发现数学上关于数字的真实，而按照程序在一有限的公理集范围内工作的计算机则因受哥德尔定理的制约，不能证明我们所能发现的真实。

不管我们建造的机器有多么复杂，它都受制于哥德尔程序，发现某一公式在这一系统中不能证明。该公式机器不能证明为真，但人运用智力却可以看出它是真的。因此，用机器来作为精神的模型仍然是

不适当的。⑧

无疑，很多人觉得，把精神的尊贵地位建立在难解的数学上，这就让人觉得难受。这是因为，平时被提出来作为非机械的精神或“灵魂”的证据的，是像爱、审美力、幽默之类的品质。总之，人们根据若干理由批驳了卢卡斯的观点。例如，霍夫斯塔特就指出，实际上人类心智发现复杂的数学真实的能力是有限的，因而一个人仍可以为一台计算机编制程序，使这台计算机能够证明一个特定的个人就数学所能发现的一切。而且，人们也很容易认识到，我们碰上了埃庇米尼底斯型的陈述，就会像计算机面对哥德尔不完备定理一样束手无策：建构涉及史密斯的关于世界的逻辑真实是可能的，但这些逻辑真实永远也不能被史密斯证明！

正如前面所强调指出的那样，意识、自由意志的印象以及自我感都涉及自指的成分，因而都可能有悖论的方面。一个人感觉到什么时（比如，感觉到一个物体），按定义讲，观察者是外在于观察的客体的。尽管他被某种感觉机制与被观察的客体联系起来，但是，在内省时，也就是说一个观察者观察他自己时，主体和客体就以一种最令人迷惑不解的方式重合起来。这就好像是观测者同时既在他自身之内，又在他自身之外。

这一奇妙的精神拓扑结构也可以用一些奇妙的图形表示出来。例如，我们可以看看有名的莫比乌斯带。把某种带子扭转一下，使它连成一闭合的环。在该带子的任何一点，似乎都有内面和外面。但是，假如你沿着已成环的带子看一遍，就会发现实际上它只有一面。局部地看，似乎有内外之分（相当于主体和客体），但从全局看则只有一面。

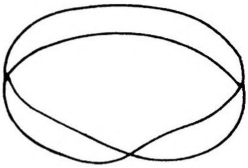


图 10　著名的莫比乌斯带。把一条带子扭转一下，使其两端连在一起，作成一个环。仔细观察就会发现，这带子现在只有一个面，也只有一条边。

关于自指的另一个启发性的图形是由霍夫斯塔特提供的，用的是他那怪圈语言：

我认为，对大脑的“显现性”现象——如想法、希望、意象、类比、最后还有意识与自由意志——进行解释的基础在于一种怪圈，也就是两个层面的相互作用，上面的层面反求于下面的层面，并影响它，同时，又被底下的层面所决定……自我一旦有能力反映它自己时，自我也就产生了。⑨

上述探求更好地理解自我的所有方案都有一个主要特点，就是承认等级层面是缠在一起的。由大脑细胞与其电化学机件构成的硬件支持着思想决策之类的软件层面，而思想、决策之类反过来与神经层面联系起来，因而修正并维持自己的存在。想把大脑与精神、肉体与灵魂分离开来的企图，是因为没有搞清这两个层面是缠在一起的（用霍夫斯塔特的话说，就是二者是“缠结为一体的等级”）。搞这样的分离是无意义的工作，因为使你之所以为你的，正是这层面的缠结。

引人注目的是，现代基督教教旨早就走向把大脑和精神看成是不可分离的整体，而且向这个方面走得相当远。现代基督教教旨强调，通过基督，整个的人得以复活，而不再像传统的观念所认为的那样，复活就是不死的独立的灵魂卸去其物质对应物，在某个场所以离体的形式继续存在。

然而，人们就精神所谈的一切并不仅限于人。科学似乎找不到证据证明人身上有特殊的神性，人们也看不到任何基本的理由来说明，为什么一台先进的电子计算机不能够在原则上与我们一样具有意识感。这当然不是说计算机有灵魂，而是说纠结层面的复杂缠绕产生了我们所认为的精神，这种复杂缠绕在各种各样的系统中也能发生。

然而，自我还有一个方面与低层面的确定性的描述相矛盾，这一个方面就是意志。所有的人都相信，在有各种各样的行动方案时，他们在某种有限的程度上能够进行选择。这种明显的引发行动的自由能够编制成程序，输入计算机吗？

霍夫斯塔特说，在原则上我们能够做到这一点。他把我们所具有的自由意志感说成是自知和不自知的微妙的平衡。霍夫斯塔特说，把适当程序的自指结合进计算机的程序里，像有它自己的自由意志一样计算机也将会行动起来。他试图将自由意志与哥德尔式的不完备说成是一回事。说任何一个能监测自己内部行动的系统都必然会有这种不完备（在第十章里，我们将更深入地探讨自由意志和决定论的问题）。

我们且来设想有一个人，通过上述的论述，相信了人的大脑是极其复杂而精巧的电化学机器，而其他类型的人造机器，如计算机之类，也可以通过程序而具有自由意志和像人一样的感情。相信了这些，是不是就贬低了人的精神？再回想一下“不过……”的圈套吧。说大脑是机器，这并不否定精神与感情的实在。精神和感情的实在指的是高一层面的东西（和蚁群、小说的情节、钢丝锯的锯齿图形、贝多芬的交响乐）。说大脑是机器，这并不一定意味着精神不过是机械过程的结果而已。若说大脑的活动具有确定性，因而自由意志便是虚幻的东西，这种说法的谬误，如同说生命的深层基础是无生命的原子作用，因而生命便是虚幻的一样。

若干科幻作家描写了具有精神的机器，其中最著名的是艾萨克·阿

西莫夫关于机器人的短篇小说，还有阿瑟·C·克拉克的长篇小说《2001年：太空奥德赛》。有些作家还进行了更为深入的分析，他们描绘了“精神移植”，试图澄清自我的定义。

例如，可以设想一下，假如你的大脑被取出，放入一个“大脑维持系统”，但仍是通过某种无线电通讯网与你的身体联系在一起，在这种情况下会发生什么事？（当然，这种事是目前可以预见的技术所完全做不到的，但在逻辑上没有任何理由说它不可能做到。）你的眼、耳以及其他的感官仍是同以往一样发挥着作用，你的身体也可以毫无阻碍的活动。在这种情况下，实际上一切仍同往常一样（可能你会觉得有点头轻！），只不过你可以俯视你的大脑。现在的问题是，你到底算是在哪里？假如你的身体乘火车出去旅行了，你的感觉就是一个人在旅行，这感觉与你的大脑仍在头颅中时是完全一样的。你肯定会感觉着你是在火车上。⑩

假如我们现在设想你的大脑被移植到另一个人的身体上，难解的问题可就多了。到底说你有了一个新的身体对呢，还是说那新的身体有了一个新的大脑对？你能够认为你自己仍是同一个人，只不过是有了一个不同的身体吗？或许你可以这么说。但要是那身体是异性的或是一动物的呢？你之所以为你，你的个性，你的能力，等等，在很大程度上是与你身体的化学和物理状况联系在一起的。假如在移植过程中你的记忆被消除了，那会发生什么事呢？那时，说那个新的人仍然是你还有什么意义吗？

假如考虑到自我的复制，就会出现新的问题。且设想你大脑的全部信息被放进某处的巨型计算机里，你原先的身体和大脑死亡了。是不是你仍然活着，活在计算机里？

把精神存入计算机的设想，使人觉得有可能将你在其他的计算机里制成多个复本。当然，已有很多文献涉及“多重人格”。所谓多重人格指的是精神紊乱。患这种紊乱症的人，其大脑左、右两半球之间的

联系被切断，导致出现一些精神状态，大致来说就是，左手不知道右手在干什么。

尽管某些这样的设想似乎很可怕，然而，这些设想也确实显现出我们有希望能够从科学的角度来理解永生，因为这些设想强调的是，精神的主要成分是信息。使我们成其为我们的，是大脑之内的模式，而不是大脑本身。贝多芬的第五交响乐在乐队停止演奏时依然存在，同样，把大脑中的信息转移到别处之后，精神可以照样存在。我们在上面考虑过如何能在原则上将精神移入计算机。但假如精神基本上就是“有组织的信息”，那么，表达信息的媒介就可以是任何东西，用不着必得是某一特定的大脑，或干脆用不着大脑。因此，我们并不是“机器中的灵”，倒更像是“电路中的信息”，而这信息是超越其表达方式的。

麦克埃用计算机语言表述了上面的观点：

假如一台正在完成某一给定程序的计算机着了火，烧毁了，我们肯定会说，这程序的具体体现完蛋了。但是，假如我们想使这同一个程序再有一个新的体现，那肯定不必把那烧毁的计算机再抢救过来，甚至也不必重复先前的机制。任何一个能动的媒介（甚至可用铅笔和纸张来完成程序）若表达了与原先的程序同样的结构、同样的关系序列，在原则上都可以体现那原来的程序。（11）

有了这个结论，那么，“程序”是否可以在另一个身体里重新运行（再世），或在一个我们觉得不属于物质宇宙的系统中重新运行（在天堂？），以及“程序”是否只是在某种意义上被“储存”起来（处于中间过度状态）？这一类问题，就成了开放性的问题了。就对时间的感觉来说，我们将看到，只有在程序正被完成的时候，如同一部交响乐正被演奏，时间之流才能被赋予任何意义。程序的存在，如同一部交响乐的存在一样，一旦被创做出来，其存在基本上就是非时间的了。

本章所说的是，认知科学方面的研究已越来越强调人与机器的心智具有相似性，这种强调对宗教具有多重的含义。一方面，认知科学方面的研究没有给传统的关于灵魂的概念留出多少地盘；另一方面，这些研究又为人死后继续存在留出了可能性。

因为精神是复杂的，所以，通常不在物理学的框架中对精神进行研究，因为我们已经看到，物理学在还原论的层面上对简单的、基本的东西处理得最好。然而，新物理学有一个重要的领域在基本层面上已被精神所侵入，使得物理学家们大为不解。那个领域被称作量子论。量子论引导我们进入了艾丽丝的奇境世界，而这世界正是对直穿过传统的宗教框架。

# 第八章　量子因素

“假如一个人不为量子论感到困惑，那他就是没有明白量子论。”

尼尔斯·玻尔

前面的两章所展开的论点都认为，精神并非是通常意义上的一个事物，不是一个有某种位置的具有一定体质的实在，然而它却是一种真实的存在，是大自然结构等级中的一个抽象的“高层面”概念。精神与肉体的关系是古老的哲学之谜，二者之间的关系如同计算中的硬件与软件之间的关系。但精神与肉体的联系要比平常的计算机软件、硬件的联系要紧密，就是说，作为软件的精神（程序）在霍夫斯塔特所谓的“缠结的等级”或“怪圈”中与作为硬件的大脑联结在一起，或交织在一起。这一幅自指的拼花图，就是意识的主要特征。

把硬件与软件、大脑与精神或物质与信息连接起来，这种想法对科学来说并不新鲜。在20世纪20年代，基础物理中发生了一场革命，震撼了科学界，并使人们头一次把注意力集中在观察者和外在世界之间的关系上。那革命性的理论称作量子论，是后来被称作新物理学的顶梁柱，它提供了迄今为止最令人信服的证据，证明意识在物质实在的本质中扮演了极重要的角色。

量子论现在已经好几十岁了，奇怪的是，它的那些令人震惊的观念花了这么长的时间才渗透到普通大众那里。不过，现在越来越多的人认识到，量子论中包含了一些令人惊讶的思想，使人得以洞见精神的本质以及外在世界的实在；要想寻求了解上帝和存在，就必须充分考虑量子革命。很多现代作家都发现，量子论中所用的概念与东方神

秘主义与禅宗用的概念很相似。但不管一个人的宗教论点是什么，总之量子因素是不可忽视的。

在开始讨论与量子论相关的问题之前，必须明确一点，这就是量子论主要是物理学的一个实用的分支。而且作为实用的东西取得了辉煌的成功。它使我们有了激光、电子显微镜、半导体和核能。它一举解释了化学键、原子以及原子核的结构，电流的传导，固体的机械性质以及热性质，以及一大堆其他的重要物理现象。量子论现已深入到大部分科研领域，至少在物理学中如此。在两代人的时间里，理科的大学生们把它当作理所当然的东西来学。当今，它在工程技术中被应用于很多实际的事上。一句话，量子论在其日常应用中是一个很实际的学科，有大量的证据证明其有效性。证据不但来自商业性的发明创造，而且也来自精微的科学实验。

尽管没有多少专业物理学家静下心去思考量子论会有怎样奇异的哲学含义，然而，量子论那确实不同寻常的性质在其发端之后不久就显现出来。量子论起因于人们试图描绘原子以及原子的成分，因而它主要是涉及微观世界的。

物理学家们早已知道，某些过程，如放射性之类，似乎是随机的，不可预测的。虽然大多数放射性原子都遵从统计学的规律，但某一具体的原子核在什么时刻衰变却是无法准确预测的。这种基本的测不准性涉及所有的原子及亚原子现象，因而，通常用于解释这些现象的观点就需要进行大大的修正。原子的测不准性在20世纪初被发现之前，人们以为所有的物质客体都严格遵守力学定律。力学定律使得行星一直在其轨道上运转，使得子弹飞向靶标。原子那时候被认为是缩微的太阳系，其内部的构件像精确的时钟一样运转。这种看法后来被证明是虚幻的。20世纪20年代，人们发现原子世界充满了含糊和混沌。像电子这样的粒子似乎根本就没有一个有意义的清晰的轨迹。在这一时刻发现它在这里，下一时刻，它又在那里，无法判定在某一时刻它在哪里。不仅是电子，所有已知的亚原子粒子，甚至是原子，我

们都不可能知道其具体的运动规律。细察之下，我们日常体验到的硬邦邦的物质化成了由幽灵一般逃逸的影像组成的大旋流。

不确定性是量子论的基本要素。不确定性直接导致了不可预测性。每一个事件都有一个原因吗？对这一问题，没有人会说没有。第三章已经说明了因果链如何被用来证明上帝的存在，证明上帝就是万事的最初原因。量子因素显然挣断了因果链，根据量子论，在没有因的情况下也可能有果。

早在20年代，人们就围绕着原子的不可预测性背后的意义展开了激烈的争论。大自然内在的本性是不是反复无常的，准许电子以及其他的粒子无规律地到处乱窜，没有节奏也没有理由？这不就是没有原因的事件吗？或者，这些粒子是些软木塞，在那看不见的微观力的海洋中摇荡？

大多数科学家在丹麦物理学家尼耳斯·玻尔的率领下，都认为原子的测不准性的确是内在于大自然的：机械运动的规则可以适用于寻常的物体，如小球之类，但到了原子那里，适用的就是轮盘赌的规则了。一位著名的大人物阿尔伯特·爱因斯坦却对此持有异议。他宣称："上帝并不掷骰子"。很多平常的系统，如股票市场和天气情况，也都是不可预测的。但之所以不可预测，是因为我们的知识不到家。假如我们完全了解有关的一切力量，我们就能够（至少在原则上能够）预测所有的来龙去脉。

玻尔爱因斯坦之争，不仅仅是细节之争。它涉及量子论这一科学的最成功的理论的整个概念结构。其核心是这样一个问题：原子是一个东西，还是一种抽象想象的构想，只是用来解释广泛的观察结果？假如原子真的是一种独立的实体，那么，它至少应当有位置，并有确定的运动。但量子论不这么看。量子论认为，原子只能二者有其一，但不能二者都有。

这就是著名的海森伯不确定性原理。海森伯是量子论的创始人之一。不确定性原理的意思是，你不能知道一个原子，或一个电子，或一个什么东西在什么位置上，同时又知道它在如何运动。你不仅不可能知道，而且，具有确定的位置和运动的原子这一概念本身就是无意义的。你可以问原子在哪里，并得到一个有意义的解答。或者，你可以问原子在如何运动，并得到一个有意义的解答。但是，“原子在哪里，它运动得多快？”这种问题是没有答案的。位置与运动（严格地说应为动量）构成了微观粒子实在性两个互不相容的方面。但是，假如原子没有位置，或没有有意义的运动，我们还有什么理由说它是个东西呢？

玻尔认为，原子的模糊世界只是在受观察时才变成具体的实在。没有观察时，原子就是一个幽灵。只是当你看它时，它才变成物质。你可以决定要看什么。想着它的位置，你就能在某一位置上看到一个原子。想看它的运动，你就可以看到以某一速度运动的原子。但你不能两者同时看到。观察所造成的实在是与观察者以及观察者所选用的测量方法分不开的。

假如你觉得玻尔的话令人摸不着头脑，矛盾，令人难以接受，那么，爱因斯坦跟你是一个观点。不管我们观察与否，世界不也确实是照样存在着吗？一切事物的发生，都有其原因，而不是因为被观察才发生的，难道不是这样吗？不错，我们的观察可能揭示出原子的实在，但怎能说我们的观察创造了原子的实在？的确，原子以及原子构成成分的行为方式似乎可能是既模糊，又不准确的，但这只是由于我们拙于探测这些精巧的东西。

微观粒子的二象性可以借助简单的电视机来加以说明。电视屏幕上的影像是由一些光脉冲产生的。当电视机后部的电子枪发射出来的电子打到荧光屏上时，就出现光脉冲。你所看到的电视画面相当清晰，原因是放射出来的电子数目极大。根据平均律，很多电子的累加结果是可以预测的。但是，任何一个特定的电子都具有内在的不可预

测性，因而也就不可能出现在荧光屏的任何一个特定位置上。这一特定的电子到达何处，构成了画面的哪一个部分，这都是不确定的。根据玻尔的哲学，从平常的枪支中射出的子弹是沿着精确的轨道奔向靶标的，但从电子枪射出的电子只是出现在靶标上。不管你瞄得多准，也不可能保证命中。“电子出现在荧光屏的X处”，这一事件不能够被认为是由电子枪或由什么别的东西造成的。因为，我们不知道为什么该电子会去X点，而不去其他的地方。于是，由该电子构成的荧光屏画面的一斑，就是一个没有原因的事件。你要是记住了这说法，下一次你看你所喜爱的电视节目时，怕是要觉得吃惊。

当然，没有谁说，电子枪与电子打到荧光屏上一事没有关系。这里只是说，电子枪并不完全决定电子能打到哪里。物理学家们并不认为打到靶标上的电子在到达靶标之前就已存在，也不认为靶标上的电子与电子枪之间有一条精确的轨迹。他们认为，离开电子枪的电子是处于一种中间过渡状态，只是有一些幽灵代表其存在。每一个幽灵都独自地探索通向荧光屏的道路，不过，只有一个电子实际显现在荧光屏上。

这些稀奇的想法怎么能证实呢？

在20世纪30年代，爱因斯坦设想了一个实验，他认为这实验能够揭露量子幽灵的欺骗，并一劳永逸地证明每一个事件都有一个不同的原因。该实验的原理是，大群的幽灵不是独自行动的，而是共同行动的。爱因斯坦说，假设一个粒子一分为二，其两半碎片可以在不受干扰的情况下作反向运动，运动到相当远处。尽管二者相距相当远了，但每一碎片都具有其同伴的印记。比如说，假如一半碎片以顺时针自转的方式飞去，那么，另一半碎片就要以逆时针自转的方式朝相反的方向飞。

幽灵理论则认为，每一碎片的自转都有一个以上的潜在可能方式。用上面的例子说就是，A有两个幽灵，一个顺时针转，一个逆时

针转。哪一个幽灵成为实在的粒子，得等到一个确定的测量或观察进行之后才能知道。同样，作相反运动的B也是由两个自转方向相反的幽灵代表的。然而，假设测量了A，使其顺时针转的幽灵成为实在，B就没有选择了，它必须使其逆时针转的幽灵升格为实在。这两个分离开来的幽灵粒子必须相互配合，以便于作用力与反作用力的定律相一致（见图11）。

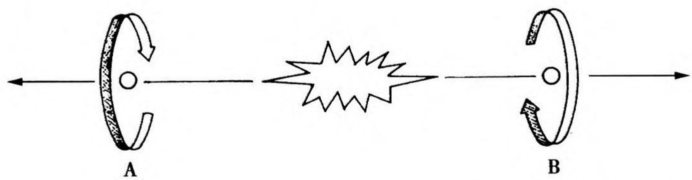



图 11　一个原子或一个亚核粒子的衰变，能产生两个旋转方向相反的粒子（例如光子），这两个粒子向相反方向运动，很可能运动到相当远的位置。

B如何能知道A选择了它的两个幽灵中的哪一个？这问题至少是令人困惑的。假如碎片A和碎片B相隔相当远，真是很难明白它们能如何通讯。而且，假如同时观察这两个碎片，那么，两者就是想传送什么信号，时间上也来不及。爱因斯坦坚持说，假如两块碎片在分离的那一刻实际上不存在（即已经按一确定的方式旋转），幽灵理论就是自相矛盾的。他还说，两块碎片作反向飞行时，它们都保持着各自的旋转方式。实际上不存在幽灵，根本没有延迟到测量发生时的选择，两碎片之间也没有那神秘的不用通讯的合作。

玻尔回答说，爱因斯坦在推理中假定了两个碎片分别都是实在的，理由是它们相距相当远。玻尔断言，人们不可能把世界看成是由许多分离的碎片构成的。在进行实际测量之前，A与B必须看作是单一的整体，即使它们相隔几光年之遥。这的确是道道地地的整体论！

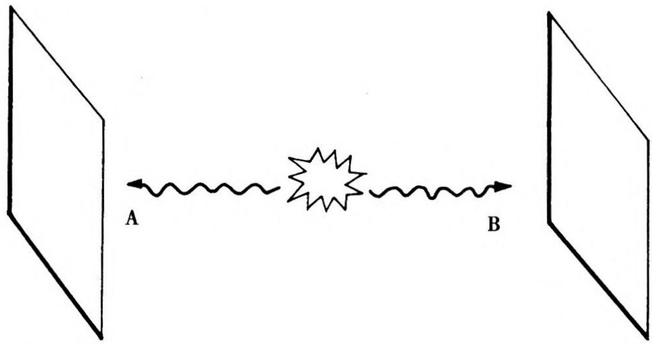



图 12 假如两个具有相关旋转和偏振的光子碰撞上平行的两块偏振材料，这两个光子就会显示出百分之百的合作：假如光子 A 被阻挡了，光子 B 也被阻挡。尽管（1）光子和起偏镜碰撞的实际结果是完全不可预测的，（2）两个光子可能相距很远，但这种合作仍然发生。

爱因斯坦对玻尔的理论提出的挑战在第二次世界大战之后才得到真正的验证。在20世纪60年代，物理学家约翰·贝尔证明了关于爱因斯坦所提出的那类实验的一个最引人注目的定律。贝尔证明，在一般情况下，如果按照爱因斯坦的看法，假定两块碎片在被观察之前实际上就已经以相当确定的状态存在，那么，分离的系统之间协作程度不能超出某个一定的极大值。量子论则预言，这个所谓的极大值是可以超出的。如今，需要的是一个实验。

技术的进步使得用实验来检验贝尔的不等式成为可能。物理学家们进行了好几个这样的实验，其中最好的一个是阿莱纳·阿斯贝及其同事1982年在巴黎大学做的。他们用一个原子同时放出的两个光子来作亚原子碎片。在每一个光子经过的通道上，他们放置了一块起偏材料。起偏材料滤出了那些其振动与起偏材料的光轴不匹配的光子。这样，只有那些与起偏材料光轴匹配的幽灵光子才会穿过起偏材料。再者，光子A与B是协作的，因为作用力与反作用力迫使它们的偏振是平行的。假如光子A被阻挡住了，光子B也要被阻挡住。

真正的测验是这样的：把两块起偏材料摆放得不成直角，那么，两个光子间的协作程度就降低，因为这两个光子的偏振现在不能够与其各自的起偏材料同时匹配了。玻尔-爱因斯坦之争在这里就可以决出个胜负。爱因斯坦的理论所预言协作程度，比玻尔的理论要低得多。

那么，实验的结果如何呢？

玻尔赢了，爱因斯坦输了。巴黎大学的实验，连同70年代所做的其他不那么精确的实验都证明，对微观世界内在的测不准性是没有什么好怀疑的。没有原因的事件、幽灵影像、只有通过观察才会出现的实在，这一切因为有了实验的证据，显然必须得到承认。

这个令人震惊的结论有什么含义呢？

很多人觉得，只要大自然只限于在微观世界里淘气，那么，就用不着为微观世界硬邦邦的实在消解了而感到很不安。在日常生活中，一把椅子仍旧是一把椅子，不是吗？

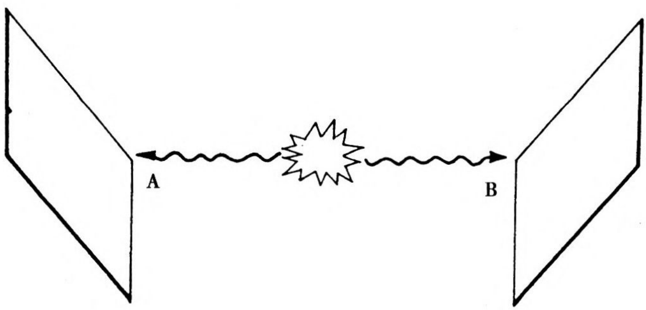



图 13　测验贝尔的不等式：假如起偏镜摆放得不成直角，A 与 B 的协作程度就降低了，有时，A 得以通过，B 却被阻挡住了。然而，却发现仍有残留的协作。那些设定（1）外在世界的独立的实在和（2）相距相当远的两个光子之间不存在神秘的逆时间通讯的理论，不能够解释光子为什么会有如此之多的残留的协作。

这得说，不完全是。

椅子是由原子构成的。原子的一大群幽灵怎么会结合成为实在的、硬邦邦的东西呢？作为观察者的人又如何呢？人有什么特殊性，使他有能力把原子的幽灵聚成坚硬的实在呢？观察者必须是人吗？猫行不行？计算机呢？

量子论是最难解又最具技术性的科目之一。我们这里的简短评论只能揭开那神秘的面纱之一角，好让读者一窥其不同寻常的概念（我在另一本书《其他的世界》中更详细地讨论了量子论）。不过，这里的素描式描述也确实说明，人们对世界的常识性看法，即把客体看作是与我们的观察无关的“在那里”确实存在的东西，这种看法在量子论面前完全站不住脚了。

量子论很多令人困惑的特点，都可以用一种奇特的“波-粒”二象性来理解。这使人想起了精神-肉体的二重性。根据“波-粒”二象性，一个像电子或光子这样的微观实体有时行为像个粒子，有时行为像一个波。是波还是粒子，要依所选择的实验来定。粒子与波是完全不同的东西。粒子是一小块浓缩的物质，而波则是无定形的运动，能够扩散和消失。一个东西怎么会既是粒子又是波呢？

这里又是并协性的问题了。精神怎么可能既是思想，又是神经兴奋呢？一部小说怎么能既是一个故事，又是一大堆字词呢？波-粒二象性是一种软件-硬件的双重性。粒子的一面是原子的硬件面——这些小球在格格地碰撞。波的一面则相当于软件，或精神，或信息，因为量子波与人们所知的任何种类的波都不同。它不是任何实体或物质构成的波，而是知识或信息的波。它是一种告诉我们就原子来说我们能了解什么的波，而不是原子本身构成的波。没有谁说原子会像波一样扩散开去。但是，能够扩散开去的是观察者对原子的了解。我们都知道犯罪率的波动，犯罪率的波不是物质的波而是或然率的波。在犯罪率的波最强的地方，最有可能出现重罪。

量子波也是一种或然率的波。它告诉你在什么地方可能有粒子，粒子具有这样或那样的属性（如旋转的方向、能量的大小之类）的可能性有多大。因此，量子波也就包含了量子因素的固有的测不准性和不可预测性。

托马斯·杨的双狭缝系统是说明波-粒二象性的冲突和二重性的最好实验。根据经典物理学长久以来的传统观点，光是一种波，是一种电磁波，是电磁场的波动。然而，在1900年，麦克斯·普朗克从数学上证明，光波在某些方面行为可以如同粒子一样——现在我们已把这些粒子叫光子。普朗克认为，光是以不可分的块状或批量的形式（量子的名词quantum[量]这个希腊词就是这么来的）传播的。爱因斯坦使普朗克的思想进一步精确化了。他指出，这些光子微粒子把原子的电子击离原子，这正是现在大家已经见惯不怪的光电管中发生的事，有些奇怪，但还不离谱。

但是，当两道光线合并起来时，最初的意料不到的事发生了。假如两个波系统重叠了，就会出现所谓的干涉效应。可以想象一下，两块石头相距几寸，落入一平静的池塘。当扩散开去的波纹重叠时，就出现了复杂的波形条纹。在一些区域，两个波运动的相位一致，于是，波动就放大了，在另一些区域，波的相位不一致，于是就相互抵消了。

要想用光来得到相同的效应，我们可以用光来照射屏幕上的两个缝，通过两个缝的光波扩散开来，就形成了干涉条纹。这很容易用照相底版显示出来。两个缝的映象并不是两个模糊的斑点，而是由明暗斑块组成的条理分明的图样，分别标明了在哪些地方两个光波是同步的，在哪些地方两个光波是不同步的（见图14）。

这一切在19世纪开始时人们就都知道了。然而，一考虑到光的微粒子性质，这就有些奇特了。每一个光子都是打到照片底版上的某一特定的点上，形成了一个小点。大群光子如同万箭齐射打到底版上，

形成了成百上亿的光点，于是就出现了底版上的影像，就像电视机荧光屏上的情况一样。任何一个单个的光子要到达哪一点，这是完全不可预言的。我们所知道的只是，光子打到底版的明亮区域的可能性很大。

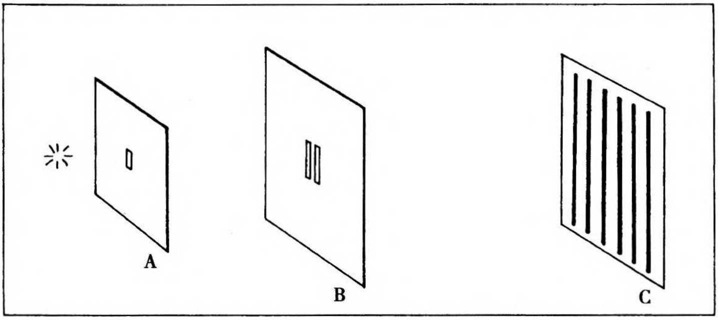



图 14　著名的杨氏双缝实验典型地揭示了光的奇妙的波-粒二象性（此实验也可用电子或其他粒子来做）。屏幕 A 上的小孔照亮了屏幕 B 上的两个狭缝。两个狭缝的影像显示在屏幕 C 上。显示出来的，并不是两道光，而是一系列明、暗条纹（即干涉条纹）。形成干涉条纹，是因为从两个狭缝发出的光波在某些位置上是同步到达屏幕 C 的，在另一些位置上则是不同步的。即使每一次只有一个光子通过实验装置，仍是照样形成同样的干涉条纹，尽管任何一个给定的光子只能经过屏幕 B 上两个裂孔中的一个，而且它也没有邻近的光子可用来为它量“步”。

然而，还不仅仅就是这些。假如我们降低照明度，使每一次只有一个光子通过实验系统，那么，只要时间足够长，由众多光子累加造成的斑点仍是组成明暗相间的干涉花纹。这里，令人难解的问题是，假如说任何一个特定的光子肯定只能通过一个缝，那么，干涉花纹则表明必定是有两个重叠的波系列，每一个缝都发出一个波系列。整个的实验也可用原子、电子、或其他的亚原子粒子来做。不管用什么粒子来做，都会出现明暗相间的条纹，表明光子、原子、电子、介子等等都同时显示出波与粒子的特征。

在20世纪20年代，玻尔为这一难题提出了一个可能的解决方法。可以把光子通过狭缝A的情况看成一可能的世界（世界A），把光子通过狭缝B的情况看成另一个可能的世界（世界B）。不知为什么，这两个世界即世界A与世界B一同出现了，叠加起来了。玻尔断言，我们不能说我们的经验世界就是A或就是B，而应当说我们的经验世界纯粹是这两个可能的世界的混合物。而且，这种混合的实在并不是两种可能性的简单相加，而是二者难以捉摸的结合：每一个世界都干涉另一个世界，形成了那有名的条纹。两个可能的世界叠加、结合在一起，颇像两张电影片同时打到一块屏幕上。

爱因斯坦是位死心塌地的怀疑论者，拒绝承认所谓混合的实在。他又提出一个经过修改的双缝实验继续与玻尔争论。在这个实验中，屏幕是可以自由移动的。他坚持认为，只要仔细观察，就可以判定光子走的是那个缝。若光子通过左边的缝，就会稍微向右偏转，那么，原则上就能够看到反冲的屏幕移向左边。假如屏幕移向右边，就说明光子通过了右边的缝。用这种方式，就可以通过实验判断是世界A还是世界B相当于实在。而且，这样一来，原先实验中光子行为明显的不确定性也就可以归结为实验技术的粗糙造成的。

玻尔对爱因斯坦提出的实验思路进行了决定性的反驳。他说，爱因斯坦这是在游戏的中间改变游戏规则。假如屏幕可以自由移动，那么，其移动也同样是受量子物理学内在的测不准性支配的。玻尔很容易地证明，准许屏幕反弹，将会毁坏照相底片上的干涉花纹，而只留下两个模糊的斑点。要么是屏幕固定，使光的波一样的性质以干涉花纹的形式显现出来。要么就让屏幕自由移动，使光子有确定的轨迹。但这么一来，光的波一样的特征就消失了，光的行为就纯粹是微粒一样的了。这里是两个不同的实验。它们不是矛盾的，而是并存的。爱因斯坦提出的实验思路并没有对原先实验中光子的路径问题进行任何说明，而原先的实验确实显示了那混合的世界。

玻尔与爱因斯坦的交锋，可以使人得出一个不同寻常的结论，那

就是，我们这些实验者以一种基本的方式参与了实在性质的形成。假如我们固定住屏幕，就可以建构出一个神秘的混合世界，在这个世界中，光子的路径是确定不了的。

1979年，在普林斯顿召开了一个纪念爱因斯坦诞辰一百周年的专题讨论会。约翰·惠勒在会上说了一番具有讽刺意味的话。他从双缝实验得出一个更让人吃惊的结论。他指出，只要对实验设备稍微进行些改造，就能够延迟选择测量方式，直到光子通过了屏幕之后再选择测量方式。这样，我们要制造一个混合世界的决定就可以在这世界出现之后再做出！惠勒说，实在的确切性质，要等到一个有意识的观察者参与之后才能确定。如此说来，可以让精神对实在进行逆时间的创造——即便是人类存在之前的实在，也可以由精神创造出来。这就是第三章中所提到的逆动因果关系。

前面的叙述表明，量子论打碎了人们的常识所珍视的关于实在性质的概念。它使得主体与客体、原因与结果之间的界限模糊了，将强烈的整体论观念引入了我们的世界观。我们已经看到，在爱因斯坦实验中，两个相距很远的粒子如何必须被看作是一个系统。我们还看到，若是不在特定的实验安排当中谈论原子的状况，甚至谈论原子这个概念是没有意义的。原子在何处以什么方式运动，这种问题是不能问的。你得首先明确你想测量什么，是位置还是运动，然后你才能得到一个有意义的答案。测量要动用大量的宏观仪器。因而，微观的实在与宏观的实在是不可分的。然而，宏观是由微观构成的，仪器是由原子构成的！这又是怪圈。

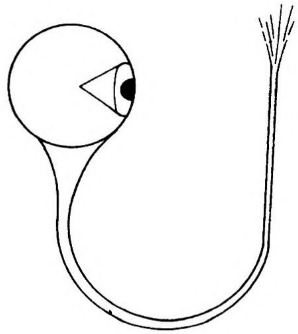


图 15　这幅象征性图画，出自约翰·惠勒。图画表示宇宙是一个自观察的系统。惠勒以令人惊讶的方式，修改了杨氏的双缝实验，表明了可以让今天的一个观察者来部分地产生遥远的过去的实在。图画中的尾部可用以表示宇宙的早期，而后来的意识对宇宙的早期进行的观察使宇宙的早期升格为具体的实在。同时，意识本身也是依存于它所形成的实在的。

大卫·玻姆这位著名的量子理论物理学家在其《整体性与暗含的秩序》一书中探讨了这些问题：

量子论所要求的关键性的描述变化就是，放弃分析的想法，不再把世界分析成相对自主的部分，分别存在但同时又相互作用。相反，现在最受强调的是不可分的整体性，在整体的世界中，观察工具与被观察的东西不是分开的。①

一句话，世界不是相互分离却相互联系的东西的集结，而是一个关系网络。玻姆这样就回应了沃纳·海森伯的话：“习惯上把世界分成主体与客体，内心世界与外部世界，肉体与灵魂，这种分法已不恰当了。”

宏观世界（即我们日常经验的世界）决定微观实在，而宏观世界又是由微观实在构成的，这个怪圈该怎么解开呢？当我们寻问在进行量子测量时实际发生了什么时，就立刻碰上了这个问题。观察者怎么会把模糊的微观世界推入具体的实在状态之中呢？

量子的“测量问题”实际上是精神-肉体或软件-硬件问题的变体。物理学家和哲学家已经与它相持了好几十年了。硬件——粒子——是用波来描述的，而波则将一个观察者观察粒子时有可能发现的关于粒子的信息变成了密码。进行观察，就使波“崩溃”成为一种具体的状态，这种具体的状态将一个确定的值赋予了被观察到的东西。

假如从头至尾全是在硬件层面上描述测量行为，就出现了悖论。可以设想一个电子正从一个靶那里扩散开去。它可以往右去，也可以往左去。你计算一下波，看看波往哪里去。波从靶那里折射开去，一部分向右扩散，一部分向左扩散，比如说，向右向左的强度相同。这就意味着，你在观察时，将会发现电子或者在左边，或者在右边。不过得要记住，在实际观察进行之前，是不可能说（也不可能有意义地讨论）电子实际在靶的哪一边的。在你进行实际的探测之前，该电子一直保持着自己的选择。两个可能的世界以一种混合的、模糊的叠加方式共存（见图16）。

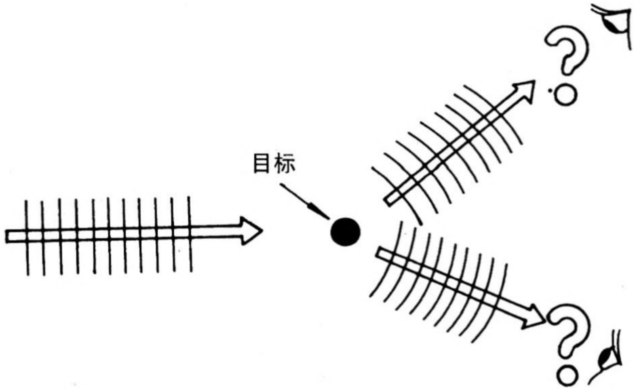



图 16　被描述为一种波的一个电子从一个靶那里弹开，产生了波纹向左向右扩散。在进行观察，判定电子向哪个方向偏转之前，必须认为是有两个幽灵世界（或幽灵电子）共存于一种非实在的混合状态中。若进行观察，其中的一个幽灵就立刻消失了，与它相关的波也就崩溃了。电子就从先前的中间过渡状态升格为单一的具体实在。现在还不知道，观察者做的什么事使电子发生了这突然的升格。是不是精神作用于物质？是否宇宙也分裂为两个平行的实体？

现在你进行观察，比如说发现电子在左边，右边的“幽灵”立刻就消失了。波突然崩溃了，倒向了靶的左边，因为现在电子已不可能是在右边了。为什么会发生这种戏剧性的崩溃呢？

为了进行观察，就必须将电子与一种外在的装置或一系列装置连接起来。这些装置被用来测试电子在哪里，并把电子存在的信号放大到宏观的层面以便记录下来。但这些连接和装置作用本身也涉及原子的机械活动（尽管涉及的原子很多），因而也是从属于量子因素的。我们可以写下一个波来代表测量装置。假设测量机器有一个指针，指针有两个位置，一个位置表示电子在左边，一个位置表示电子在右边。那么，把电子和测量装置组成的整个系统看成是一个量子大系统，就迫使我们得出一个结论：那个左右不定的电子的混合性质转移到测量装置的指针上了。测量装置必须进入一种量子中间过渡状态，

而不能显示其指针指向左还是指向右。这样，测量行为似乎就把那讨厌的量子世界放大到实验室的尺度上了。

数学家约翰·冯·纽曼对这一难题进行了探讨，用了一个简单的数学模型证明，将电子与测量装置连接起来，结果的确促使了电子选择在左边还是在右边，但这只能是将电子混合的非实在状态转移到测量装置的指针上。冯·纽曼还指出，假如测量装置再与另一个装置连接起来，这一个装置能够显示出第一个装置的测量结果，那么，第一个装置的指针因而也就被促使做出选择。现在，进入中间过渡状态的是第二个装置。于是，可以有一连串的机器互相测量，得出有意义的“非此即彼”的结果，但冯·纽曼机器链总是有一台最后的机器必得处于一种非实在的状态之中。

薛定谔的一个著名实验突出地表明了量子论中这难解的问题。在这个实验中，有一个放大装置被用来引发一种毒剂的释放，该毒剂能杀死猫。于是，左-右指针的二重性转换成了活-死猫的二重性。假如一只猫要被描述为一个量子系统，那么，我们就被迫得出结论：在一个人或一个什么物去观察猫之前，猫处于一种两可的“活-死”状态，而这状态似乎是荒谬的。

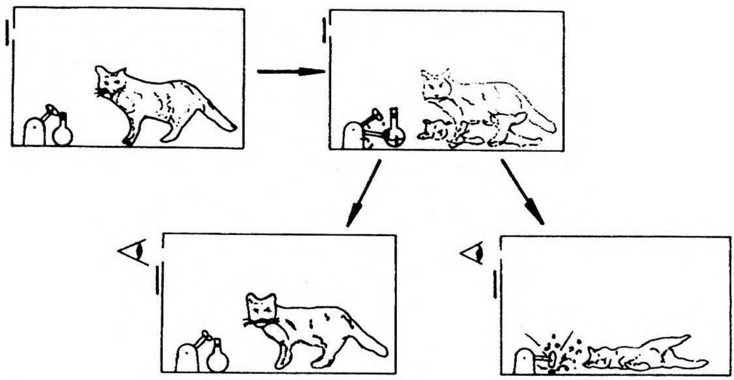


图 17　薛定谔猫的悲惨故事。一个量子过程可以 50：50 的概率使杀猫剂释放出来。量子论认为，整个系统必得进入幽灵般的混合状态，在进行观察之前，猫就是处于这种死–活的混合状态。一旦进行了观察，就会发现那猫不是活的就是死的。这种想象实验突出了量子论中观察行为的不可思议的含义。

假设用一个人来代替猫，这代替猫的人能够体验到一种活-死状态吗？当然不能。那么这就是说，量子力学到了人类观察者这里就失效了吗？难道说，冯·纽曼的机器链到了人的意识那里就到头了吗？有一位杰出的量子物理学家，尤金·威格纳，实际上就提出了这个令人激动的主张。威格纳认为，量子系统的信息进入了观察者的精神，使得量子波崩溃了，并且突然地将那种混合的、两可的幽灵状态转变成为泾渭分明的具体的实在状态。于是，当实验者观看装置的指针时，就使得指针做出偏向左边还是偏向右边的决定。因而，靠着机器链的传导，也使得电子打定了主意要往哪边去。

假定我们承认威格纳的说法，那就是又承认了早先的二元论：精神是独立于物质的但又与物质处于同一个层面的实体，精神可以作用于物质，使之发生显然违反物理定律的运动。对此，威格纳倒不以为意。他说，“意识对（大脑的）物理-化学状态有影响吗？或者说，人的肉体违反那种通过研究无生命的物质得来的物理定律吗？对此问题

的传统回答是，‘不’，肉体影响精神，但精神却不影响肉体。然而，至少可以举出两个理由来支持相反的观点。”②威格纳举出的一个理由是作用与反作用的定律。假如肉体作用于精神，那么，精神也反作用于肉体。另一个理由就是前面提到的由之而来的量子测量问题的解答。

必须承认，同意威格纳说法的物理学家很少。尽管有些物理学家借用量子的精神作用于物质的思路，主张某些超自然的现象如意念致动和遥距意念弯物之类是可能的。（“假如精神可以启动神经元，为什么就不能弯曲调羹呢？”）

威格纳的论点，从头到尾有强烈的层面混淆的味道。试图借助于软件（精神）来讨论硬件（四处奔忙的电子）的运行，便是掉进了二元论的陷阱。然而，这里的问题更为棘手，因为硬件和软件在量子论中是死死地缠结在一起的（如，波-粒二象性）。不管威格纳的论点正确性如何，其论点确实表明，精神-肉体问题的解决可能是与量子测量问题的解决密切相关的，不管最后结果如何。

另有人试用的解开量子测量难题的方法，或许比威格纳求助于精神还来得离奇。只要人们所处理的是有限的物理系统，冯·纽曼的机器链就可以延长。你总是可以说，你所观察到的所有东西都是实在的，因为有一个更大的系统，使你通过“测量”或“观察”所看到的东西崩溃成为实在。但是，近年来，物理学家们注意到了量子宇宙学，也就是关于整个宇宙的量子论。从定义上讲，没有任何东西可以处在宇宙之外，来使整个的宇宙崩溃成为具体的存在。（或许，上帝是例外？）在这一层面上，宇宙似乎就处于一种中间过渡状态，或宇宙两可状态。假如没有一个威格纳的精神来使宇宙一体化，宇宙似乎就注定要这么不死不活的延续下去，只是一堆幽灵，是多种可能的实在多重混合又叠加，其中哪一个实在也不是实际的实在。那么，为什么我们还是感觉到一个单一的、具体的实在呢？

有人提出了一个大胆的设想来正面解决这一问题，这设想就是平行宇宙的理论。休·埃弗列特1957年创立了这一理论，后来得到了现在奥斯汀得克萨斯大学任教的布莱希·德·威特的支持。该理论认为，所有可能的量子世界都是同样地实在，而且是平行存在的。一旦进行一次测量来测定，比如说，猫是活是死，那么，宇宙就一分为二了，一个包含着活猫，另一个包含着死猫。两个世界都同样实在，二者也都包含着人类观察者。然而，每一个世界的居住者都只能察觉到他们自己所在的宇宙。

宇宙因为一个电子的古怪行径就分岔，由一变为二，这种不寻常的概念或许是人们的常识难以接受的。但是，进一步仔细察看，就会发现这理论颇站得住脚。当宇宙分岔时，我们的精神也随之分了岔，成为两个副本，一个世界住一个。每一个精神副本都以为自己是独一无二的。那些认为感觉不到自己分了岔的人，应该想一想他们也感觉不到地球绕太阳转的运动。随着所有的原子、所有的亚原子粒子的四处跳跃，分岔的事也就一次又一次地发生。宇宙在每一秒钟之内都被复制无数次。用不着进行一次实际的测量以便使这种复制发生。只要一个单独的微观粒子以某种方式与一宏观的系统互相作用就够了。用德·威特的话来说：

每一个恒星，每一个星系、每一个宇宙的遥远的角落发生的每一个量子跃迁都使我们这里的世界分岔，变成成千上万个它们的副本……这就是带有某种报复的两可状态。③

为恢复实在所付出的代价是实在的多重性——数目庞大而且不断增加的平行的宇宙沿着各自的演化分支在发散开去。

其他的世界是什么样子的呢？我们能不能去呢？可不可以在其他世界里找到关于不明飞行物或百慕大三角区人、船失踪之谜的解释？对研究不明飞行物专家们来说，遗憾的是，埃弗列特的理论在这一系列问题上态度明确。平行的宇宙一旦分离之后，从物理上讲就完全分

家了，互不相通了。要想使它们再聚在一起，就得将测量的过程倒回去，也就相当于让时间倒转。这很像是一个原子一个原子地把一个打碎的鸡蛋再恢复原样。

但是，这些其他的世界又在哪里呢？从某种意义上说，那些与我们所在的世界很相像的世界都靠我们很近。然而，这些其他的世界是完全不可进入的。不管我们在我们的时空中走得多远，也到不了这些其他的世界。本书的读者离他自己成千上万的副本不过一寸之遥，但这一寸却不是用我们所感知的空间所测量得出的。

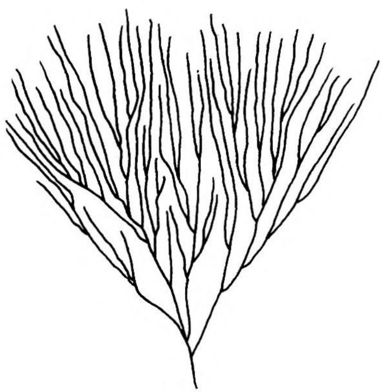


图 18　为了回避活-死猫以及其他的量子两可非实在，埃弗列特提出，量子系统的不确定性产生了一种多枝状实在，在这种实在中，宇宙不断分岔，变为众多的“平行宇宙”，这些平行宇宙在物理上是不连通的，但都是同样实在的。观察者的精神也随着宇宙的分岔而分成无数的副本。

各个世界分岔分得越远，它们彼此间的差异也就越大。以某种细微的方式与我们的世界分离的那些世界如在双缝实验中的光子的路径，假如不仔细看，就看不出它们与我们的世界的差异。然而，在其他与我们的差异较大的世界中，猫的数目可能与我们的世界不同。在某些世界里，希特勒没有出现过，约翰·肯尼迪还活着。还有些世界差

异大得不得了，尤其是那些在时间开始后不久就分岔的世界。实际上，一切可能发生的事（尽管不是一切可以在想象中发生的事）都在这多枝状的实在的某一个枝上发生着。

所有可能的世界同时存在，这就引起了一个令人感兴趣的问题，即为什么在有人读这本书的这个世界是现在这样的世界，而不是在不同的分支上的其他的世界？显然，本书的读者不可能在所有的或大多数其他的世界中存在，因为在那些世界中，没有适合于生命发生的条件（在第十二章我们还要再回来讨论这个问题）。

很多人提出，量子论为我们了解自由意志开辟了道路，因为量子论是以非常基本的方式与精神相联系的。先前人们用决定论的观点看待宇宙，认为在宇宙中我们所做的一切都是在我们出生以前很久由宇宙的结构决定好了的。现在，这种观点似乎被量子因素扫除了。这难道是说，自由意志现在活得还好吗？要想正式地讨论这个问题，我们必须首先较为深入地探究一下时间之谜。

# 第九章　时间

“假如不预先假定过去和将来有区别，经验这个词就不会有意义。”

卡尔·冯·魏泽克

“但我总是听到在我背后，有翼的时间之车在迅速驶近。”

安得鲁·马维尔

两场大革命促成了新物理学的诞生，一场是量子论，一场是相对论。后者几乎完全是爱因斯坦一人的成就。相对论是一个关于空间、时间和运动的理论。其影响同量子论一样，既深刻又令人迷惑，并对人们抱有的关于宇宙本质的很多观念构成了挑战。相对论对时间的处理，是有史以来对人们的传统观念最有挑战性的，因为时间是世界上各大宗教长期以来所热切关心的一个问题。

就我们对世界的经验而言，时间是如此基本的东西，以至谁若想摆弄摆弄它，便会遭到人们的反抗，遭到极大的怀疑。每个星期，我都收到一些业余科学家们的来稿。这些人一心想找出爱因斯坦的错误，企图再将传统的、常识性的时间概念复原，尽管过去将近80年以来，相对论一直是成功的，还没有任何一个实验证明按相对论做出的完美无缺的预言有什么错误。

我们对个人同一性的看法，即对自我、对灵魂的看法，是与记忆、与延续的经历密切相关的。只是说在此刻“我存在”还不够。是一个人，就意味着经验的连续，还有将经验连接起来的某些特征，如记

忆。这个问题具有很强的宗教意味，也是一个让人容易动感情的问题。很可能就是因为这个缘故，新物理学的那些看法才受到了抵制，同时科学家和科学门外汉们，都深深地被相对论的那些让人不知所措的推论所吸引。

爱因斯坦的所谓狭义相对论发表于1905年。狭义相对论起源于爱因斯坦试图消除物体的运动和电磁扰动的传播之间明显的矛盾。光信号的行为尤其显得违反人们信奉已久的原理，即一切匀速运动都是完全相对的。我们在这里不必去讨论那些技术性细节。反正结果是，爱因斯坦重建了相对性原理，使之即使在涉及光信号的情况下也能成立，但做到这一点是有代价的。

狭义相对论的第一个受害者是人们对时间的信仰——时间是绝对的，普遍的。爱因斯坦证明，时间实际上是有弹性的，可以被运动伸长或压缩。每一个观察者都带着他自己的时间尺度，而他的时间尺度在一般情况下是与别人的不一样的。在我们自己的参照系中，时间从来也不会显得有什么异常，但相对于另一个以与我们不同的方式运动的观察者而言，我们的时间就可能是被扭曲了，与他们的时间不同步了。

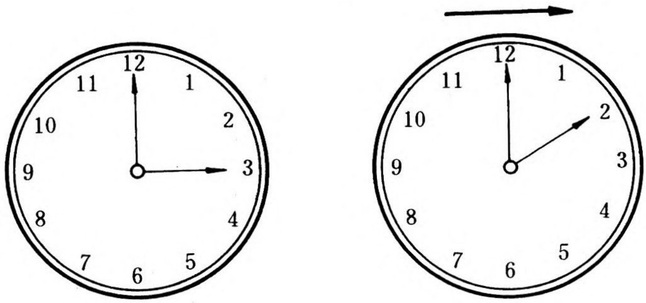



图 19 时间膨胀效应现在对物理学家们来说已是司空见惯了。可以用高速运动的灵敏的原子钟或具有已知的衰变率的亚原子粒子显示出时间膨胀。运动的钟相对于不运动的钟而言，其走时要慢一些。这个现象就导致了著名的“双生子效应——一个宇航员高速航行若干年后返回地球，变得比他留在地球上的孪生兄弟年轻了。

时间尺度发生这样奇妙的错乱，使得我们有可能进行一种时间旅行。从某种意义上说，我们都是时间中的旅行者，都在奔向将来，但时间的弹性使得一些人能比另一些人早一些到达将来。高速运动使你能够减慢你自己的计时器的运转，好像是让别人冲在时间的前面。用这种办法，就有可能比静坐不动更快地到达某一遥远的时刻。从原则上讲，我们可以在几小时之内就到达2000年。然而，要想得到可观的时间弯曲，就必须有每秒几万英里的高速。现有的火箭速度，只能使精确的原子钟显示出些微的时间膨胀。时间膨胀的关键是光速。随着我们接近光速，时间弯曲也逐步升级。相对论禁止任何人超过光速，因为超过光速，就会出现时间倒转的情况。

可以用高速的亚原子粒子使时间戏剧性地缩短。μ介子在巨大的回旋加速器中被加速到接近光速，可享有十几倍于它静止时的寿命（它静止时，大约在1微秒之内就会衰变）。

相对论也同样使空间受到了重大损害，使空间也成为具有弹性的了。当时间被伸长时，空间就被缩短。假如你坐在列车上，列车驰过

车站，从你的参照系来看（你的参照系是与站台上的搬运工的参照系相对的），车站上的钟走得要稍微慢一点。作为补偿，站台在你看来也显得短了一些。当然，这些事我们从未注意过，因为在常规的速度下，时钟走时和站台长短的变化太小，不过这些很小的变化很容易用灵敏的器具测量出来。空间和时间的这种共同的扭曲可以看作是空间（收缩了）变成了时间（伸长了）。1秒钟的时间相当于很大很大的空间——准确地说是相当于299000公里。

科幻小说里常有这种时间扭曲的花招，但时间扭曲的确不是虚构的。这种扭曲真的会发生。有一个奇特的现象，即所谓的双生子效应，就说明了这种扭曲。一个孪生子以接近光速的高速飞向我们邻近的一个恒星。他那呆在家里的孪生兄弟等了10年，终于等到他返回地球。火箭着陆之后，他的孪生兄弟发现在这10年里，他只长了1岁。高速使他只过了1年的时间，而在他的1年里，地球上已过了10年。

爱因斯坦将其狭义相对论作了进一步的推广，使之包括了引力效应。于是就有了广义相对论。在广义相对论中，引力不是一种力，而是时空几何中的一种扭曲。按广义相对论来看，时空并不是服从通常“平坦的”几何学规则的，而是弯曲的，产生时间弯曲和空间弯曲的。

我们在第二章里说过，现代的工具十分灵敏，连地球引力的时间弯曲都可以用火箭里的钟探测出来。在太空中，时间走得确实要快些，因为在那里，地球的引力比较弱。

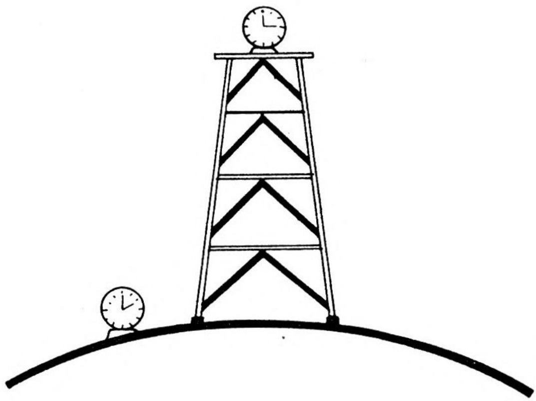


图 20　引力使时间减慢，这在地球上也可以用实验证明。相对于塔脚下的钟而言，塔顶上的钟走时快。

引力越强，时间弯曲也就越明显。现已知道，在有的恒星上，引力十分强大，以致那里的时间相对于我们要慢百分之几。实际上，这些恒星正处于某种临界值的边缘，一过了临界值，时间弯曲就开始加倍增长。假如这样的恒星再大几倍的话，时间弯曲就会升级，最后，在引力的某一临界值上，时间就会停下来。从地球上看出，这恒星的表面就是冻结住了，没有任何活动了。不过，我们是看不到这种奇异的时间停止现象的，因为我们借以看到它的光线也冻结在那里了，这恒星发出的光的频率被移到了光谱的可见区域之外。这恒星看上去是黑的。

理论告诉我们，处于这种状态的恒星不可能保持原状，而会在它自己强大的引力压迫下在1微秒之内坍缩成为一个时空奇点，在空间留下一个空洞，即黑洞。原先恒星的时间弯曲仍然在空洞的空间中留有痕迹。

因此，黑洞就代表着通向永恒的近路。在这种极端的情况下，火箭上的那位孪生子不但可以快一些到达将来，而且能够在一瞬间到达时间的终点！他一旦进入黑洞，黑洞之外的一切永恒从他相对固定

的“现在”来看，就会立刻成为过去了。因而，他一进入黑洞，就会被锁闭在一种时间弯曲之中，不能再返归外面的宇宙了，因为外面的宇宙都已经过去了。就宇宙的其他部分而言，他的确是处于时间的终点之外了。他要想从黑洞中出来，就必须在进入黑洞之前就出来。这是荒谬的，说明他不可能从黑洞中逃出来。黑洞的引力毫不留情地死死抓住这倒霉的宇航员，把他拖向奇点，到了奇点，在1微秒之后，他就到达了时间和湮没的边缘；奇点就标志着通向“无空间”和“无时间”的单程旅程的终点。奇点是自然宇宙终结之“地”。

相对论在我们的时间观念中引发的革命，可用如下的话做出最好的概括。就是说，从前，时间被认为是绝对的，不变的，普遍的，是独立于物体和观察者的。现在，时间被认为是能动的。时间能够伸长收缩，弯曲，甚至可以在奇点处停止。钟表的走时不是绝对的，而是与运动状态和观察者的引力状况相对的。

把时间从普遍性的紧身衣中解救出来，使得每一个观察者的时间都能够自由而独立地运行，这就迫使我们放弃一些长期抱有的假设。比如说，现已不可能就选择何时为“现在”达成一致意见了。在双生子的实验中，火箭上的那一位在向外飞的途中或许会想“现在我那孪生兄弟在地球上干什么呢？”但这两位孪生兄弟相对时间尺度的错乱意味着，火箭参照系中的“现在”与地球上的人所判定的“现在”是完全不同的时刻。没有普遍的“现在”。假如在不同的地点发生了两件事A和B，一个观察者认为A与B同时发生，另一个观察者就会认为A先于B，而第三个观察者就可能认为B先发生。

两个事件发生的时间顺序在不同的观察者看来竟会不一样，这似乎是不可思议的。靶标在发枪之前就会碎吗？谢天谢地，没有这种事。否则该多出多少伤亡。要想使事件A与B发生的顺序难以确定，A与B就得在足够短的时间内发生，使光在该时间内来不及从A跑到B。在相对论中，光信号是一切的规则，而光信号尤其禁止任何信号跑得比光信号快。假如光不能快到把A与B联系起来，就没有任何东西能把

二者联系起来，因此，A与B就不可能以任何方式相互影响。二者之间没有因果联系：将A与B的时间顺序调换过来，就不会造成因果颠倒。

世上没有普遍的现在，这一事实不可避免地使那种把时间整齐地划分为过去、现在和将来的做法遭到腰斩。过去、现在和将来这些术语在我们周围的场所中可能有意义，但不能适用于别的地方。诸如“现在火星上正在发生什么事？”这一类的问题指的是该行星上的某一特定的时刻。但正如我们已经看到的那样，一个乘火箭掠过地球的太空旅行者在同一时刻问同样的问题，指的就是火星上的另一个时刻了。实际上，在地球附近的一个观察者根据其运动的情况的不同，其可能的“现在”要有好几秒钟那么长。观察者离被观察对象越远，“现在”的范围也就越大。对一个遥远的类星体来说，“现在”可能是一段几十亿年的时间。即便是在地球上漫步这样的运动，也会使类星体上的“现时刻”变为几千年！

因为人们长久以来是如此相信只有现在“确实存在”，所以，抛弃那种把时间整齐地划分为过去、现在、将来的做法，就是人类思想历程中意义深远的一大进步。人们通常不假思索地认为，将来是尚未形成的，很可能还没有确定下来；过去则是过去了的，虽留在记忆中，却是泼出去的水，没法子了。人们希望相信，过去与将来都不存在。“每一次”似乎只有一瞬的实在发生。相对论使得这一切观念成了无意义的东西。过去、现在和将来必定是同样实在的，因为一个人的过去是另一个人的现在，再一个人的将来。

一个物理学家对时间的看法很受他对相对论的了解的影响，因而很可能显得与常人有相当的差异，尽管物理学家自己对这事并不怎么在意。物理学家并不认为时间是由发生的事件构成的一个序列。相反，他们认为，过去和将来的一切都在那儿，时间在任何一个给定的时刻都向过去和将来两个方向延伸，就像是空间在任何一个给定的位置延伸一样。事实上，这里把时间和空间相比较还算不得什么，因为时间和空间在相对论中已经变得交织在一起密不可分了，两者合成为

物理学家们所谓的“时空连续体”了。

我们对时间的心理感觉与物理学家们的时间模型的差别是如此之大，以至连许多物理学家也觉得这中间是不是缺了点什么重要的东西。爱丁顿曾经说过，我们的精神有一种“后门”，时间除了通过常规的感官途径和实验室仪器的途径进入我们的精神，还走“后门”。我们对时间的感知，比起对空间方位或物质的感知来，是更为基本的。对时间的感知是一种内在的体验，而不是一种肉体的体验。我们尤其感觉得到时间的流逝这种感知是如此明显，以至构成了我们感觉的最基本的一面。这一面是一个有力的背景，我们的一切思想，一切活动都是在这一背景的衬托之下被感觉到的。

很多科学家在寻找那神秘的时间之流时被搞得晕头转向。所有的物理学家都承认，宇宙中存在着过去-将来不对称，这种不对称是由热力学第二定律造成的。但要是仔细检查一下该定律，这种不对称似乎就没有了。

这个难题可以简单地说明如下。假设在一间封闭的房间里，把一个香水瓶的瓶塞拔去。不一会儿，香水就挥发出来，弥漫整个房间，房间里所有的人都闻得到。香水由液体变为香气，从有序到无序的转变是不可逆的。不管过多长时间，散布开来的香气分子也不可能自动地返回香水瓶，规规矩矩地再变回为香水液体。香气分子的挥发，散布，就是过去和将来不对称的一个典型例子。假如我们在电影上看到，香气分子返回香水瓶，我们立刻就可以知道，这电影片子是倒放的。因为香气分子一从瓶子里冒出来，就不可能再回去的。

然而，这里却有一个问题。香气分子的挥发，散布，是几十亿分子撞击的结果。空气分子的热运动一刻也不停，把香气分子到处乱撞，使之移来移去，最后香气分子终于不可收拾地与空气混合到了一起。然而，任何一次特定的分子碰撞都完全是可逆的。两个分子相互接近，碰撞、倒退。在这事上没有什么时间不对称。相反的过程仍是

相互接近，碰撞，倒退。

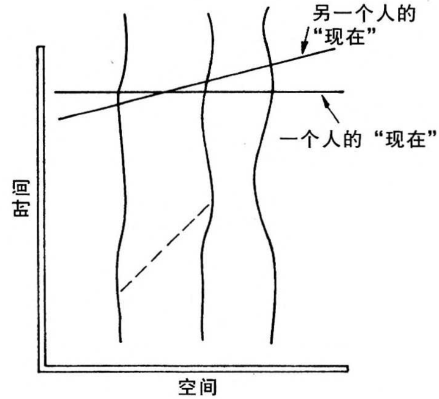



图 21　物理学家们并不认为时间是在流逝。他们认为，时间作为“时空连续体”的一部分摆在那儿。在本图中，时空连续体的四维结构被绘成二维的面，其中两个空间维度被省略去了。面上的一点代表一个“事件”。图上的曲线代表运动物体的路径，虚线代表在两物体之间传播的光信号。图中的水平线代表一个观察者在某一时刻所看到的整个空间的一个切面。另一个观察者因为以不同的方式运动，所以，他所看到的整个空间的切面就是一个斜面。于是，为了理解世界，就必须有时间的纵向延伸。不存在一个代表单一的、共同的“现在”的普遍“切面”。因此，把宇宙切分成过去、现在，将来就办不到了。

对称的分子碰撞怎么会产生不对称的过去和将来，这个由时间之矢构成的谜牵动了很多杰出的物理学家的想象。路德维希·玻尔兹曼在19世纪后期首先阐述了这个问题，但直到今天，大家仍是对此议论纷纷。有些科学家断言，存在着一奇异的非物质的性质，即时间流，正是这时间流造成了时间之矢。他们断言，分子运动在通常的情况下是不能在时间上打下过去-将来不对称的印记的。因此，要想形成时间不对称，就必须有额外的成分，即时间流。有人甚至到量子过程中或宇宙的膨胀中去寻找时间流的起源。相信有时间流在很多方面就像是相信有生命力一样，都很让人怀疑。

相信有时间流是个错误，它错在忽视了这一事实：时间不对称如同生命一样，是一整体概念，因而是不能把整体的性质归结为单个分子的性质的。在分子水平上的对称和在宏观尺度上的不对称，这二者之间没有什么矛盾。它们纯属两个不同的描述层面。于是，人们就想，时间怕是并非真的在“流动”，一切都是我们的头脑在作怪。

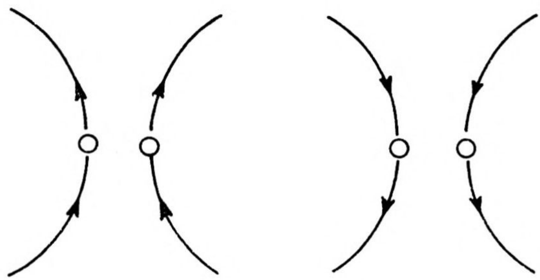


图 22　假如我们在分子水平上观察物质，就会发现时间不对称在世界的起源上是个不解之谜。任何两个分子之间碰撞的过程都是完全可逆的，没有任何从过去到将来的倾向。

假如我们想在感觉中找到时间流的起源，我们就会碰到在理解自我时所碰到的那些混乱和悖论。我们就很难不得出这样的印象：时间流的问题和自我的问题是密切相关的。只有在时间的流水中，我们才能感觉到自我。霍夫斯塔特曾写道，“自指的旋涡”造成了我们所说的意识和自我意识；而我则坚决相信，推动心理上的时间流的也正是那同一个旋涡。正是因为这一点，我才坚持认为只有在我们解开了时间之谜后，才能解开精神之谜。

在艺术和文学中，到处都可以发现关于时间的天真想象。时间被描绘成箭，是河流，是飞车，时间在前进。人们常说，我们意识到的“现在”在不断地向前推移，由过去移向将来，这样，2000年也会最终变成“现在”；同理，你读这个句子的时刻现在也已经过去，成了历史。有时，现在被认为是固定的，时间本身是流动的，就像是河水流

过河边的一个观察者。时间的这些形象是与我们对自由意志的感情不可分割的。将来似乎尚未形成，因而就能够在它尚未到来时，用我们的行动来使它成形。这一切是不是胡说八道呢？

假如说，上段里所说的关于时间的形象都是对的，那么，就立刻会出现一大堆问题。1983年，在一个物理学家和一个怀疑论者之间或许有一场如下的对话：

怀疑论者：我刚刚看到爱因斯坦的这一句话："你必须承认，主观的时间将着重点放在现在，而主观的时间并没有客观的意义……过去、现在、将来之间的区别，只是一种幻觉，不管人们怎么坚持这种区别也没有用。"爱因斯坦肯定是发了疯吧？

物理学家：绝对没发疯。在外在的世界中，根本就不存在什么过去、现在、将来。你能用仪器测定出现在吗？不能。现在是个纯心理概念。

怀疑论者：哦，你是在开玩笑吧？谁不知道，将来还没有发生，而过去已经过去了，我们都记得过去发生过了。你怎么能把昨天和明天或今天混为一谈呢？

物理学家：当然，你得把那么多天理出个次序来，但我所反对的是你用的那些标签。怕是连你自己都承认，明天永远不会来。

怀疑论者：别玩文字游戏好不好？明天怎么不会来？只不过来了之后，我们把它叫作今天就是了。

物理学家：一点不错。每一天在当天都叫作今天。每一时刻在被人感觉到的时候都叫"现在"。过去和将来的划分是语言混乱的结果。我来帮你澄清这个混乱吧。时间的每一个时刻都可以有一个确定的日期。例如1997年10月3日下午2点。记日期的系统是任意的，但一旦约定俗成地确定下记日期的系统，任何一个事件或时刻的日期就确定

下来了。所有的事件，我们都给它一个日期标签，这样，我们在描述世界上的一切事的时候，就可以不用过去、现在、将来这些模模糊糊的东西了。

怀疑论者：但是，1997年的确是将来。1997年还没来到呢。你所说的日期系统忽略了时间的一个关键方面，即时间的流动。

物理学家：你说“1997年的确是将来”是什么意思？1997年是1998年的过去。

怀疑论者：但现在不是1998年。

物理学家：现在？

怀疑论者：不错，就是现在。

物理学家：什么时候是现在？每一个时刻在我们感觉到它的时候，都是现在。

怀疑论者：此刻是现在。我说的现在是指此刻。

物理学家：你说的是1983年的现在？

怀疑论者：就算是吧。

物理学家：不是1998年的现在？

怀疑论者：不是。

物理学家：好，那你说的意思就是，1997年是1983年的将来，但却是1998年的过去。这我并不否认。我的日期系统所描述的恰恰也正是这么回事。所以你瞧，你根本用不着说什么过去将来之类的东西。

怀疑论者：不用区分过去和将来？这真是荒谬。1997年还没来，

这一点你要承认吧？

物理学家：自然要承认。你所说的意思是，我们的谈话发生在1997年之前。让我再重复一遍：我并不否认众多事件有一个顺次的序列，它们之间有先后或过去将来的关系。但我所否认的是过去、现在、将来的存在。显然，并没有一个现在，你和我在一生中都会感觉到很多“现在”。

一些事件在另一些事件的过去或将来。但事件只是在那儿存在，并非是一个接一个地发生。

怀疑论者：有些物理学家说，过去和将来同现在并排存在，过去、现在、将来都在那儿，只是我们顺次地一个挨一个地遇见过去、现在、将来的事件罢了。你说的话跟这些物理学家是一个意思吗？

物理学家：我们实际上并不是“遇见”事件。我们只是感觉到所有的我们意识到的事件。从时间的角度来看，并没有一大堆事件等在那儿让我们去悄悄地接近。众多事件就在那儿，是人的头脑把它们联系起来了。你刚才说的那番话，意思就好像是今天的精神被用一种什么方法给向前挪，好去碰上明天的事件。其实，你的精神是延伸在时间之中的。明天的精神状态反映明天的事件，今天的精神状态反映今天的。

怀疑论者：难道是我的意识从今天向明天移动吗？

物理学家：不！你的精神今天和明天都是有意识的。没有什么东西向前向后，或向左向右移动。

怀疑论者：但我明明感觉到时间在流逝。

物理学家：且慢1分钟。真对不起。说“且慢1分钟”这话意思是让你等等，我先说几句。但这话听起来倒好像是时间是你手里的东西，你能拖住它似的。言归正传，首先，你说你的精神在时间上是往前挪

的，后来，你又说时间本身在向前挪。你到底是说的哪个呢？

怀疑论者：我把时间看成是一条流动的河，将来的事件朝我冲过来。我既可以把我的意识看成是固定的，时间流过我的意识，从将来流向过去；我也可以把时间看成是固定的，我的意识从过去移向将来。我想，这两种说法都是一回事。运动是相对的嘛。

物理学家：你这运动是幻想的东西！我问你，时间怎么个动法？假如说时间在运动，那它肯定就有速度。请问：时间的速度多大？每天1天？这不是绕口令吗？1天就是1天的1天。

怀疑论者：要是时间不动，事物怎么会发生变化？

物理学家：这好说。物体在时间里活动，变化于是就发生了。可时间并没动。我小时候常纳闷："为什么这会儿是现在？为什么另外一个时候不是现在？"我长大了才知道，这样的问题是无意义的。"这会儿"可以是任何一会儿。

怀疑论者：我倒是认为，你小时候提的那个问题很有道理。不管怎么说，为何这就是1983年？你总得说出个所以然吧？

物理学家：为何什么就是1983年？

怀疑论者：哦，为何现在就是1983年？

物理学家：你问的问题很像是问："为何我是我、不是别人？从定义上说，我就是我自己，就是那个问问题的人。显然，在1983年，我们就把1983年看作是"现在"。在哪年，就把哪年看成现在。你其实该问："为什么我活在1983年，而不是活在，比如说，活在公元前5000年？"你也可以问："为什么我们在1983年而不是在1998年进行这场谈话？"你看，这么问，就用不着动用过去、现在、将来之类的概念。

怀疑论者：我还是不明白。我们日常的思维活动，如语言的时态结构，希望、恐惧、信仰，等等，这一切几乎都是根植于对过去、现在、将来的基本区分上的。我害怕死亡，因为我尚未碰到它，我不知道死亡之后是怎么回事。但是，我虽不知道我出生之前是怎么回事，可我却不害怕。我们不可能害怕过去。我可以再说一遍，过去的事是改变不了的了。靠着我们的记忆，我们知道过去发生了什么事。但我们不知道将来。然而我们相信将来是未定的，我们的行动可以改变将来。至于现在嘛，可以这么说，所谓现在，就是我们跟外在世界发生接触的一瞬间。在这一瞬间里，我们的精神可以命令我们的肉体去行动。拜伦曾写道："行动吧，要在活的现在行动"。他这句话，就把我的意思，全说明白了。

物理学家：你说的大部分都对，但你仍用不着用一个移动的现在来说明问题。当然，过去和将来之间存在着不对称。这不对称不光是记忆之类的感觉所感觉到的，在外在世界中也确实存在着过去和将来的不对称。例如，热力学第二定律表明，一切系统会越变越乱，越无序。我们大脑之外的其他系统也具有累积的记录和"记忆"。你可以想想月球上那么多的环形山。那都是对过去事件的记录，不是对将来事件的记录。你刚才说的那番话，意思不过是说，后来的大脑要比先前的大脑储存了更多的信息。可是，我们接着犯了一个错误，把这至浅至明的事实译成了混乱又模糊的话："我们记得过去，却不记得将来。"其实我们都知道，"过去"是个无意义的词儿。到了1998年，我们将能记得1997年，尽管1997年是1983年的将来。只要你按日期说就能说明白，用不着时态，或用时间之流，现在之类的玩意儿。

怀疑论者：可你刚才自己就用了个将来时"将能记得"。

物理学家：我其实完全可以说，"我1998年的大脑记录关于1997年的事件。但1997年是1983年的前头，因此，我1983年的大脑没记录1997年的事"。瞧，我这里就用不着使用过去和将来的概念。

怀疑论者：可是，对将来的恐惧、自由意志和不可预测之类又该怎么解释呢？假如将来早已存在，这岂不成了彻头彻尾的决定论了吗？什么都早定好了，改不了了，自由意志不就成了冒牌货了吗？

物理学家：将来并非“早已”存在。你的话本身就有矛盾。因为你的话等于说将来的“事件与那些先于它们的事件同时存在”，按“先于”一词的定义来看，这显然是无意义的。至于说到不可预测性，那也是有实际的局限的。不错，由于世界太复杂，我们只能预测某些简单的事件，如日食之类。但可预测性与决定论并不是一回事。你是把你的认识论同你的形而上学搞混了。世界将来的状况都可以是被在先的事件决定了的，但实际上照样还是不能预测世界的将来状况。

怀疑论者：但是，难道将来是被决定了的吗？对不起，我不该说将来。我是说，难道一切事件都完全是被在先的事件决定了的吗？

物理学家：实际上不是这样。比如，量子论就说明，在原子水平上，事件都是自发地发生的，没有什么完全的前因。

怀疑论者：这就是说，不存在将来！我们可以把它变来变去！

物理学家：不管有没有我们现在的行动，将来就是将来。物理学家把时空看成像是展开的一张地图，时间在图的一边展开。事件在图上是一些点。有些事件由因果关系与先前的事件联系在一起，其余的事件，如放射性原子核的衰变，则被标作是“自发的”。不管有没有因果联系，一切都在那儿了。所以，我所说的没有过去、现在、将来之分并没有涉及自由意志或决定论。这完全是另外一个问题，这是一个布满了混淆的领域。

怀疑论者：你仍是没有给我解释为什么我感觉到时间的流逝。

物理学家：我不是神经病学家。不过，你之所以有那样的感觉或许跟短期记忆过程有关。

怀疑论者：你是说，时间的流逝完全是主观的问题，是一种幻觉？

物理学家：求助于感觉来把一些物理性质归因于外在世界，这是不智之举。我问你，你感觉过眩晕没有？

怀疑论者：当然有过那样的感觉。

物理学家：可你并没有把你的眩晕归因于宇宙的旋转，尽管你的确感觉到世界在打转儿。

怀疑论者：这倒是。因为眩晕时觉得世界在打转儿，显然是一种幻觉。

物理学家：所以我说，时间的流逝，就像眩晕时感觉到的空间在打转儿，不过是一种时间性的眩晕而已。时间流逝这种幻觉之所以有了一个虚假的真实外表，是我们混乱的语言造成的。语言里的时态结构和那些关于过去、现在、将来的无意义的词语使得时间流逝的幻觉像是实有其事。

怀疑论者：请详细说说好不好？

物理学家：现在不行。我没时间了……

从这样的谈话里我们能得出什么结论来呢？无疑，我们在安排日常事务时是大大倚仗过去、现在、将来这些概念的。我们从不怀疑时间真是在流逝。即便是物理学家，要是头脑里一放松分析推理这根弦，就会很快地在言谈思维中像常人一样认为时间是流逝的。然而，必须承认，我们越是仔细检查过去、现在、将来这些概念，这些概念似乎就越是变得难以把握，晦涩不明，我们的陈述最后不是同义反复，就是无意义。在物理学的世界里，物理学家不需要时间的流逝或现在之类的东西。实际上，相对论干脆就排除了所有的观察者有一个共同的、普适的现在的可能性。假如过去、现在、将来这些概念确有

意义（很多哲学家如麦克塔戈特否认这些概念有意义①），那么，其意义似乎要归属于心理学而不是物理学。

于是，这就引起了一个令人迷惑的神学问题。上帝有没有时间流逝的感觉？

基督徒认为，上帝是永恒的。然而，“永恒”一词有两个颇为不同的意义。其较为简单的意义是永存的，没有开始和终结的，无限延续的。可是，有人对基督徒的这种看法大不以为然。处在时间之中的上帝是要有变化的。但是，使上帝发生变化的原因又是什么呢？假如上帝是万物万事的原因（如同第三章里的关于上帝存在的宇宙论证明所说的那样），那么，谈论最终原因本身变化岂不是无意义吗？

在前面的几章里，我们已经看到，时间并非仅仅存在而已，其本身也是自然宇宙的一部分，时间是“有伸缩性的”。根据明确的、依赖于物质行为的数学法则，时间能伸也能缩。同时，时间与空间也是紧密相连的。时间和空间一起表达了引力场的运作。简言之，时间像物质一样，在所有的细节上，参与了自然过程。时间并非是神圣不变的，而是能改变的，用物理方式，甚至用人工。因而，处在时间之中的上帝在某种意义上说，也是受制于自然宇宙的。实际上，时间很可能在将来的某一阶段停止存在（第十五章将会讨论这个问题）。这样，上帝的地位显然就不保险了。很清楚，假如上帝是受制于时间物理的，那他就不可能是全能的。假如他没有创造时间，也就不能认为他是宇宙的创造者。事实上，因为时间和空间是不可分离的，那么，上帝没有创造时间，也就是没有创造空间。但正如我们已经说过的那样，时空一旦存在，完全是自然的活动就会自动地造成宇宙中的物质和秩序。因此，很多人认为，只要上帝创造出时间（严格地说应当是时空），其他的一切就用不着上帝来创造了。

于是，我们就得出了“永恒”这个词的另一个意义，即“时间之外的”。上帝是在时间之外的，这种概念至少自奥古斯丁以来就有了。

我们在第三章讲过，奥古斯丁认为，上帝创造了时间。奥古斯丁这种看法获得了很多基督教神学家的支持。圣安赛尔姆将其看法作了如此的表述：“你（上帝）不存在于昨天、今天、明天，而是存在于时间之外。”②

时间之外的上帝就没有上面提到的那些麻烦问题，但却有第三章上讨论过的那些缺点。时间之外的上帝不可能是一个人格的上帝，不可能思想、说话、有感觉、筹划，因为这一切都是时间性的活动。很难想象时间之外的上帝怎么会在时间之内行动（尽管有人说这不是不可能的）。我们也说过，自我的存在与时间流逝的感觉是多么密切地相联系的。时间之外的上帝算不得我们所知的任何意义上的一个人格（Person）。这一类的疑惑促使若干现代神学家拒斥上帝永恒的观点。保罗·蒂里希写道：“假如我们说上帝是活的，我们就是肯定上帝包含时间性，因而就与时间的变化有关系。”③卡尔·巴斯也持有相同的观点：“假如上帝的时间性不完全，基督教启示的内容就不成形。”④

时间的物理对人们相信上帝是全知的这一信仰也具有很有意思的含义。假如上帝是在时间之外的，就不能说他会思想，因为思想是一种时间性的活动。但是，时间之外的一个存在会有知识吗？获得知识显然是要时间的，但知识本身却不需要时间，假如所知的东西本身不随时间变化的话。假如上帝知道今天的每一个原子的位置，那么，上帝的知识明天就会变化。上帝若是具有不受时间限制的知识，就必须知道贯串时间的一切事件。

因而，要把上帝所有传统的属性调和起来就有了一个严重而根本的困难。现代的物理学发现了时间的变易性，就在上帝的全能和上帝的人格存在之间打入了楔子。现在已很难说上帝既能全知又具有人格了。

# 第十章　自由意志与决定论

“一切都是确定的；将来如同过去，我们都可以看得着。”

彼埃尔·德·拉普拉斯

当牛顿发现他的力学定律时，很多人都以为，自由意志的概念这下子算是寿终正寝了。牛顿的理论认为，宇宙像是一个巨大的钟表，钟表的弦正严格地按预先定好的方式放松，最后松到不能再松的地步。据认为，每一个原子的行踪都是事先定好的，在时间开始时就确定下来了。人类不过是这巨大的宇宙机制之中的附属机器，一进入这机制之中就逃脱不了。后来，出现了新物理学，随之也有了时空的相对性和量子的不确定性。于是，选择自由和决定论的问题又重新热闹起来。

构成新物理学基础的两个理论是相对论和量子论。这两个理论似乎彼此怀有根本性的敌意。一方面，量子论认为，观察者在形成物理实在的本质时具有十分重要的作用；我们说过，很多物理学家认为，有确凿的证据证明，不存在所谓的“客观实在”。这似乎使人类获得了一种独一无二的能力，使人能够以一种牛顿时代梦想不到的方式影响自然宇宙的结构。另一方面，相对论则破除了普遍时间的概念，也破除了绝对的过去、现在、将来的概念，把将来描绘成在某种意义上说早就存在的东西，因而将我们借助量子论获得的胜利就这么给打发掉了。假如说将来早已存在，那么，是不是说我们无力改变将来了呢？

按照牛顿从前的理论来看，每一个原子都沿着一个独一无二的轨道运动，其轨道是由作用于该原子的所有的力决定的。而作用于该原

子的力则是由其他的原子决定的，以此类推。牛顿力学认为，在了解某一时刻的情况的基础上，原则上可以精确地预测将来会发生的一切。世上存在着严格的因果网络，一切现象，从一个分子的极微小的跳动到一个星系的爆炸，都是很早以前就连细节都定下来了的。正是因为这种力学观点，使得彼埃尔·德·拉普拉斯（1749-1827）宣布，假如谁知道某一时刻宇宙中的每一个粒子的位置和运动情况，他便掌握了所有的必要信息，可以计算出整个宇宙的过去和将来的历史。

然而，这种拉普拉斯式的计算却并不像看上去的那样简单。首先，大脑能否计算出它自己将来的状况（即便是在原则上计算）就成问题。麦克埃曾经指出，对每一个人来说，完全的自我预测是不可能的，即使是在牛顿式的机械宇宙中也不可能①。可以设想，有一个超级科学家能够窥探你的大脑，精确地计算出在将来的某一个场合你将做什么；然而，这在逻辑上并不排除你有某种意义上的自由意志。原因是，尽管这位超级科学家或许预测得正确，但他却不能告诉你（在事前）他的预测，否则，他就会搞乱他的计算。例如，假如他跟你说，“没错，你要拍手”，你的大脑状况必然因而发生变化，变得与他跟你说这话之前不同。之所以会出现这种情况，是因为你获得了来自他的新信息。这时，你就不能相信他的预测了，因为他的预测是按照发生变化之前的大脑状况做出的。因而，不可能有任何预测能够准确地预测你会正确地想见到你未来的行为。于是，麦克埃认为，不管你的行为在那假想里掌握着预测的超级科学家看来是多么的可以准确预测，多么的不可避免，但对你自己来说在逻辑上却仍是不可预测的，因此也就至少是保留了通常所谓的自由意志。

还有一个问题是，宇宙按照牛顿力学是否是可以预测的？近来，在对力学系统进行数学描述方面取得了一些进展。这些进展表明，某些力能造成一些系统在其演化过程中产生强烈的不稳定，因而，可预测性便是一个无意义的概念。在“通常的”系统中，初始状态的微小变化只能造成系统行为的微小变化；然而，那些超敏感的系统则会因初

始状态的极其微小的差异而产生全然不同的演化。而且，现代宇宙学的发现也表明，我们的宇宙在空间中具有膨胀的视界，每一天都有新的扰动和影响从视界之外的区域进入我们的宇宙。因为那些区域自时间起始以来与我们这部分宇宙从未有过任何因果性的交往，所以，我们即便是在原则上也不可能知道这些新来的影响是什么。

对完全可预测性的最重要的驳难来自量子因素。按照量子论的基本原则来看，大自然在本质上是不可预测的。海森伯著名的不确定性原理告诉我们，在亚原子系统运行过程中总是有一种无法消除的不确定性。在微观世界中，事件的发生并没有明确的原因。

决定论的垮台是不是与相对论矛盾呢？相对论认为，不存在普遍的现在，宇宙的整个过去和将来是一个不可分割的整体。世界是四维的，三个维是空间，一个维是时间，一切事件都在四维世界里，不会有什么将来“发生”或“展开”。

实际上，决定论的垮台与相对论并不矛盾。决定论说的是，一切事件都完全是由一个在先的原因决定的。决定论并没有涉及事件是否在那儿的问题。毕竟，将来就是将来的样子，不管它是不是由先前的事件决定的。相对论的四维图景则只是使我们不能以绝对的方式把时空切割成普遍的时刻。两个事件在不同的地方是否“同时”发生，这得看我们的运动状态如何。一个观察者或许会认为它们是同时发生的，另一个观察者则可能认为二者是相继发生的。因此，我们必须把宇宙看成是既在时间中延伸，也在空间中延伸。但相对论并未说时间的延伸是否包括宇宙中发生的事件之间的严格的因果关系。因此，尽管过去、现在、将来似乎没有客观的意义，然而相对论却并未说人不可能用在先的行动决定在后的事件（要记住，先后的次序是时间的一个客观性质，而过去和将来却不是）。

然而，目前尚不清楚的是，非确定性的宇宙是否是确立自由意志所需要的。实际上，决定论者认为，只有在确定性的宇宙中才谈得上

自由意志。毕竟，具有意志自由的人是一个能在自然世界中运用因果律来做出某些行为的人。而在一个非确定性的宇宙中，事件的发生是没有原因的。假如你的行为不是你造成的，你还能说你的行为是你干的吗？自由意志的辩护者认为，一个人的行为是由其人品、偏好、个性决定的。

我们不妨设想，一个本分的人突然犯了罪。非决定论者可能会说，“这纯是自发的事件，没有前因。这个人是不能责备的。”决定论者则会认为，这个人该为其罪行承担责任，同时，还会认为通过教育、劝导、心理疗法、药物等可以使他将来不犯罪。大部分宗教思想所传达的一个中心信息就是：我们能够改好。但是，我们将来的品行所能改好的程度是由我们在先的决心和行动决定的。在这里我们应当认识到，决定论并不是说事件的发生与我们的行动无关。有些事件之所以发生，是因为我们决定了它们的发生。

千万不可把决定论与宿命论混为一谈。宿命论说的是，将来的事件是我们完全不能控制的。宿命论者会说，“一切事其实早就由星宿定下来了，将有的事必会有。”假如一个身处枪林弹雨之中的士兵心想，“要是我气数已尽，不管怎么防备也免不了一死”，于是就在战场上无所顾忌，毫不躲避，那么，这个士兵就是一个宿命论者。某些东方宗教就有宿命论的味道，而且很多人也会时时跌入宿命论中，尤其是在涉及世界大事的时候，“那些大事我是无能为力的”。这倒是不假。普通的人的确不能制止世界大战，也不能阻止一颗大流星毁掉一个城市。然而，在日常生活中，我们不停地以数不清的小手段来影响事件的结局。毕竟没有人会认真地说：“我会不会让汽车撞死，这是命中注定的事，所以过马路的时候我不必费心去看左右来往的车辆。”

尽管决定论跟宿命论不是一回事，但我们对它还是放心不下。无怪乎很多人在得知量子因素显然推翻了决定论之后，不禁长长地大松了一口气。的确，我们对自由的要求，是要求我们所决定的事真的可

以因我们的决定而发生。但是，在一个完全确定的宇宙中，决定本身也是早就决定了的。在这样的宇宙中，尽管我们可以随意而行，但我们想做的事却是我们所控制不了的。道理是这样的：当你决定喝茶不喝咖啡时，使你做出这个决定的是环境的影响（比如，茶相对便宜），生理因素（比如，咖啡的刺激性较强，你受不了），文化的因素（茶是传统饮料），等等。决定论认为，所有的决定，所有的奇想，都是早就决定了的。假如事情果真如此，那么，不管你在决定喝茶还是喝咖啡的时候觉得多么自由，实际上，你的选择在你出生的那一刻就定了，很可能定得还要早。在一个完全确定的宇宙中，一切都在宇宙创生的时刻就定下来了。这岂不是使我们一点自由也没有了吗？

现在的问题是，很难判定我们到底要的是哪样的自由。有人说，在咖啡和茶之间做出选择的真正自由意味着，假如导致一种选择的环境再次出现时（即宇宙中的一切都跟当初做选择时一模一样，包括你的大脑状态也跟当时一样，因为你的大脑也是宇宙的一部分），你很可能在这第二次做出不同的选择。只有环境相同时选择不同，才称得上是自由。这样的自由显然与决定论大相径庭。但是，这种终极的自由怎样才能得到验证呢？宇宙怎样能照原样再来一次呢？假如自由真是这样的，那么，自由的存在必定纯是个信仰。

或许，自由意味着另一回事。也许就是麦克埃所说的不可预测性？你将要做的事是被你所不能控制的因素决定的，但你永远也不能知道，即使在原则上也不能知道你将做什么。这难道就足以满足对自由意志的要求吗？

另一种自由观认为，某些（不是所有的）事件是有原因的，但我们不能在自然宇宙中找到其原因。具体地说，这种观点认为，我们的精神是外在于物质的世界的（这是二元论哲学），但我们的精神能够以某种方式进入物质的世界，能够影响发生的事件。于是，仅就物质世界而言，并不是所有的事件都是能被决定的，因为精神并不是物质

世界的一部分。不过，人们仍可以问，使精神做出其决定的原因是什么？假如其原因来源于物质世界（某些原因显然如此），那么，我们便是又回到决定论那里去了，将非物质的精神提出来便成了放空炮。但是，假如某些原因是非物质性的，我们能因而更自由一些吗？假如我们控制不了非物质性的原因，那么，我们就跟控制不了物质性原因一样，也谈不上什么真正的自由。假如我们能够控制使我们做出自己的决定的原因，那么，又是什么决定了我们如何进行控制的呢？是更多的外在影响（物质的或非物质的）还是我们？“我之所以做出这种选择，是因为我使我自己使我自己使我自己……”，这得“使”到什么时候才到头？我们是不是都必须跌入无穷的倒退？我们能不能说，这因果链条的第一环是自动的，不需要外部的原因？这种自动因的概念，即没有原因的原因是有什么意义的吗？

到目前为止，我们一直对非决定论避而不谈。大多数物理学家会说，决定论与自由意志论之间的争论是无关宏旨的，因为我们知道量子因素已证明决定论是不成立的。但我们在此必须小心谨慎。量子的影响很可能太小，不会对神经元层面上的大脑行为产生多大的影响，但假如真有影响的话，我们肯定就不会有自由意志了，我们肯定会精神崩溃。量子起伏使一个通常不该兴奋起来的神经元兴奋起来（或使通常应该保持兴奋的神经元兴奋消退），这肯定会被看作对正常的大脑运作的干扰。假如一些电极被植入你的大脑，由一个外人随机地通电或断电，你一定会认为这种形式的干扰削减了你的自由：这个外人“接管了”或至少说是妨碍了你的大脑的运作。你大脑中量子的随机活动难道不是跟“噪音”一样吗？你决定要抬起你的胳膊，有关的神经元也按正确的次序兴奋起来了，可量子起伏打乱了信号，你的胳膊没动，腿倒动了。这难道就是自由？非决定论的致命问题就在这里：因为你的行动不是由你或由什么别的东西决定的，所以，你可能不能够将你的行动置于你的控制之下。

然而，有人仍旧摆脱不了那种虚幻的印象，以为量子因素的确带

来一些自由的希望。不错，只要我们神经元的兴奋序列一旦开始，我们就不希望它被打断。但是，有人会说，量子作用也正是在初始阶段才具有重要性。我们可以设想一个神经元已准备好兴奋，只差原子水平的那极微小的扰动来使它兴奋起来。量子论认为，该神经元是否兴奋，是有确定的几率的。实际结果如何是不确定的。精神（或灵魂）就是在这里起了作用。它（下意识地）说："电子向右运动！"或下了诸如此类的命令，于是，神经元兴奋起来。在这里，精神作用于物质，但却没有违反物理定律，因为神经元无论如何要兴奋起来的可能性显然存在。精神只是触动了一下可能性的天平，使得神经元真的兴奋起来罢了。

然而不幸的是，并没有任何证据证明大脑真有这么微妙的平衡，假如真有，外来的电磁扰动就可能搞乱大脑的作用。而且，精神作用于物质的说法也有我们上面说过的问题——一开始使精神命令电子向右运动的原因是什么？反对用二元论解释精神-肉体问题的人也会强烈反对这种说法，因为他们坚持认为，精神并不是一种能作用于大脑的物质。假如把大脑看成是一种软件，代表大脑的电化学结构，那么，再谈论精神作用于大脑就是又一次混淆层面了。这样的谈论之无意义，就如同把一本小说的出版归因于小说中的一个人物，或把计算机中的一个电路开关的开通说成是程序使它开通的。

上面所说的一切，实际上并没有真正抓住量子论的中心悖论——即精神在决定实在的过程中所扮演的独特的角色。我们前面说过，观察的行为使幽灵般叠加在一起的潜在的实在，凝聚成单一的具体的实在。在没有观察的情况下，一个原子本身是不能做出选择的。我们必须观察，然后才会有某种具体的结果出现。通过选择测量原子的位置或测量原子的运动，你能够决定是创造一个在一个特定位置的原子，还是创造有一个特定速度的原子。这一事实便证实了你的精神在某种意义的确进入了物质世界，且不管这种进入的实质是什么。但现在我们可以再一次问：为什么你决定测量原子的位置而不测量原子的运

动？这种建构实在的自由，比起那种我们已司空见惯的自由，即用移动物体的方式（比如，用接触）影响外部世界的自由，难道是更有力的吗？

如今，很多物理学家倾向于接受埃弗列特的所谓多宇宙量子论的解释。埃弗列特的观点（其观点已在第三章简短地讨论过）对自由意志具有奇特的含义。埃弗列特认为，每一个可能的世界实际上都是实在的，同时，所有其他可能的世界也都平行地共存。这些重叠的世界延伸开来供人类选择。可以设想你面临一个选择——要茶还是要咖啡？埃弗列特的解释认为，你一旦进行了选择，宇宙立刻就分为两个岔，在一个分岔里你有茶，在另一个分岔里你有咖啡。这样，你就什么都有！

多宇宙理论似乎是可以满足上面讨论过的选择自由的终极标准。当宇宙分岔时，导致每一种结局的条件在所有的方面都是完全相等的，因为这些条件实际上是在同一个宇宙里。然而，人却做出了两种不同的选择（我们在前面讲过，没有人能够直接证实多宇宙理论，因为所有的人都必定局限于分了岔的宇宙当中的一个分岔里）。然而，这理论虽满足了选择自由的标准，却似乎付出了过大的代价。假如你不得不进行所有的可能的选择，那你还有什么自由可言呢？这自由似乎是过了头，让它自己的成功给毁了。你想选择的是茶或是咖啡，是二者择一，不是两样都要。

但是，赞成多宇宙理论的人会说，“你这里的你是什么意思”？那个有了茶的“你”和那个有咖啡的“你”不是一个人，他们住在不同的宇宙里。假如要想把他们区别开来的话，那么可以说，这两个我们漫不经心地称作“你”的人在感觉经验方面是不同的，比如，这两个“你”对饮料的口味不同。这两个人不可能是同一个人。因而，假如你得到选择的机会，你实际上不会既有茶又有咖啡的。不管我们讨论的是哪个“你”，反正那个你做出了选择。那么，按照多宇宙理论的观点来看，说你选择了茶而不要咖啡，不过是给“你”下了一个定义而已。

说“我选择了茶”不过是说“我是喝茶的人”而已。因此，虽说面临选择的是一个单个的你，结果却涉及两个人，不是一个。按埃弗列特的理论来看，自我在不断地分成无数个相近的复本（这对传统的那种灵魂独一无二的概念会有什么含义，探讨起来会很有趣的）。

犯罪要承担责任、责难与自由意志的关系，有很多这方面的文献。假如自由意志是虚幻的，那为什么有人要因其行为而受责难呢？再者，假如一切事件都是早就注定了的，那么，我们每一个人就逃不脱我们出生之前就已定下的定数。在一个埃弗列特式的多重宇宙中，一个罪犯可不可以为自己辩护说，按照量子论的定律，他那多重的自我中至少有一个成员得要犯下罪行？然而，我们还是先避开这一是非之地，先询问一下上帝在一个决定论式的宇宙中的地位如何吧。我们一旦把上帝放在这种宇宙中，就立刻碰上了一大堆难题。

上帝能够施行自由意志做出决定吗？

假如人有自由意志，上帝肯定也有吗？假如有，那么，前面所讲的涉及自由的各种问题也要困扰上帝了。而且，就无限的、全能的上帝而言，我们也有很多理解不了的难题。假如上帝有一个宇宙计划，而这计划又作为其意志的一部分得到执行，为什么上帝不干脆创造一个决定论式的宇宙，使其计划目标成为不可避免的东西？或者，在创造出宇宙的同时就使其计划得到实现，这岂不更省事？假如宇宙是非决定性的，那么，上帝也不能决定或预测结果会是什么，这岂不是说他的能力是有限的吗？

或许有人可能争辩说，假如上帝愿意，他便可以自由地放弃他的一部分能力。他可以赋予我们自由意志，使我们能够逆他的计划而行，假如我们愿意这么做的话。他也可以赋予原子以量子因素，使他创造宇宙的过程变成一个宇宙规模的或然率游戏。但是，一个真正全能者是否能放弃一部分能力，这倒是一个逻辑难题。

全能意义上的自由与人类所享有的那种自由是很不一样的。你可以自由地选择茶或咖啡，但只有在茶和咖啡存在的情况下，你才能有这份自由。你没有做成你想做的任何事的自由，比如，徒手游过大西洋，或把月亮变成血的自由。人类的能力是有限的，只有一小部分愿望有可能得到实现。而全能的上帝却没有能力的限制，他能够自由地做成他想做的事。

全能这一概念也引起了一些令神学家难堪的问题。上帝能自由地杜绝恶吗？假如上帝是全能的，那就是能。但假如他能，那为什么没有做到这一点呢？大卫·休谟提出了一个对神学来说具有毁灭性的论点：假如世上的恶出自上帝的意志，那么，上帝便不是仁慈的；假如恶是违反上帝的意志的，那么，上帝就不是全能的。上帝不可能既全能又仁慈（就像大多数宗教所宣扬的那样）。

对大卫·休谟的论点，有人提出一个反驳：恶完全是由人的行为造成的；因为上帝给了我们自由，我们就可以自由地做恶事，从而破坏上帝的计划。然而，这里仍有一个问题：假如上帝能够自由无碍地预先阻止我们作恶，那么，他若是没有做到这一点，岂不是也必须承担一份责任吗？假如一个做父母的让他那任性的孩子胡作非为，在四邻中打砸抢，我们通常认为那做父母的也要承担一份责任。我们是不是因而必须得出一个这样的结论：恶（或许就一个有限的量而言）是上帝计划的组成部分？或者说，上帝不能自由地做到预先阻止我们逆他的意志而行？

假如我们按照基督教的信条进行思想，相信上帝是超越时间的，那么，我们还会碰上一大堆新的难题，因为选择自由这一概念从本质上说是一个时间性的概念。假如选择不是在一个具体的、特定的时刻进行，而是没有时间限制的，那么，选择还有什么意义呢？假如上帝早已预知了将来，那么，上帝为宇宙制订计划以及我们参与他的计划还能有什么意义呢？一个无限的上帝知道在所有的地方现在正在发生什么事。但我们已知，不存在一个普遍的现在，所以，假如上帝的知

识是延伸在空间中的，就必定也在时间中延伸。于是，我们就可以得出这样的结论：基督教的永恒的上帝具有选择自由是无意义的。但是，我们能够相信人类具有人类的创造者所没有的能力吗？我们似乎是被迫得出一个悖论性的结论，即选择自由实际上是我们所受的一种限制，就是说，是我们在知晓将来上的无能。上帝一旦摆脱了现在的牢笼就不需要自由意志了。

这些问题似乎是无法解答的。新物理学无疑为解开自由意志和决定论这一对长期悬而未决的谜提供了一个新线索，但并未把谜解开。量子论虽动摇了决定论的基础，但在涉及自由时也给它自己出了一大堆难题，而其中最难的难题是多重实在的可能性。毫无疑问，随着我们将来对时间了解的加深，这些有关我们的存在的基本问题必会出现新线索。

# 第十一章　物质的基本结构

通过寻找越来越小的物质单位，我们并不能找到基本的物质单位，或曰不可分割的物质单位，但我们却的确碰上了一个点，在这一点上，分割是没有意义的。

沃纳·海森伯

现今的寻求建立统一场理论的努力真是愚不可及。

I·M·辛格

科学之所以能成为科学，只是因为我们生活于其中的宇宙是一个井然有序的宇宙，这宇宙符合质朴的数学定律。科学家的工作，就是研究、讲述大自然的井然有序，并将其有序分门别类，而不是对大自然的秩序的起源提出疑问。但神学家们长期以来一直认为，物质世界存在的秩序是上帝存在的证据。假如神学家的看法是对的，那么，科学和神学也就是目的一致地显示上帝的工作了。实际上，一直有人认为，西方科学的出现，是由基督教—犹太教的传统促成的，因为这种传统强调，上帝有目的地把宇宙组织起来，而这种组织可以借助理性的科学研究被人们看出来。这是一种哲学，这种哲学的精义似乎可以用斯蒂芬·黑尔斯（1677～1761）下面的话来表示：

既然我们确信全知的造物主在创造万物的时候遵守了最为严格的数、重量、尺度的比例，所以，若想洞见进入我们的观察范围之内的那些被造物的本质，最有可能成功的方式必定是从数、重量、尺度入手。

宇宙之井然有序似乎是自明的。不管我们把目光投向何方，从遥远辽阔的星系，到原子的极幽深处，我们都能看到规律性以及精妙的组织。我们所看到的物质和能量的分布并不是混乱无序的，相反，它们是按照从简单到复杂的有序的结构安排的，从原子到分子，到晶体、生物，到行星系、星团等，莫不是井井有条，按部就班。而且，物质系统的行为也不是偶然的、随机的，而是有章法、成系统的。科学家们面对大自然难以捉摸的美和精妙时，常常感到一种敬畏和惊奇。

将不同种类的有序区分开来是有用的。首先，有一种质朴的有序，如我们在太阳系中或在钟摆的摆动中所看到的那种规律性。还有一种复杂的有序，如木星旋转的大气中气体的排列，或一个生物的复杂组织。这种区别是还原论对整体论的又一个例子。还原论试图揭示出复杂组织中的简单成分，整体论则注意于整体的复杂性。复杂的有序在很多人看来有一种目的成分。在复杂的有序中，一个系统所有的组成部分和谐地组织在一起，合作着去达到一个特定的目的。在这一章里，我们将看看质朴的有序，看看基本物理最新的发现如何证实了数学规律性在控制着大自然的那些重要过程。在下一章里，我们要回过头来探究复杂的有序。

康德提出，为了理解世界，人的精神不可避免地要将秩序强加于世界。但我认为很多科学家对他的观点不以为然。比如说，康德根本就不懂原子或原子核的结构，他哪能知道，后世的人们对原子进行研究，显示出原子也具有太阳系组织的那种数学规律性。这一发现的确是惊人的，跟我们所选择的观察世界的方式没有任何关系。而且，我们还将看到，亚核物质也是服从某些质朴有力的对称原理的。人们很难相信，在某些基本力的作用过程中，左右对称是无意义的，只不过是对人类精神的赞美。

人们在传统上一直遵循着科学的还原论来揭示大自然中的质朴有序：把复杂的系统分解成较为简单的部分，再分别对这些部分加以研

究。一切物质都是由少数基本单位（即最初的“原子”）组成的，这种观念起源于古希腊。但是，只是在本世纪，技术才有了足够的进展，使我们能够详细地研究、了解原子的作用。这方面最早的发现之一，主要是卢瑟福勋爵在本世纪初做出的。这个发现就是，原子根本就不是基本的粒子，而是由其内部的构件合成的结构。原子质量大部分集中于小小的原子核，原子核只有一厘米的一万亿分之一大小。核的周围包围着由较轻的粒子（即电子）构成的云，电子云延伸的距离达一厘米的一千万分之一。因此，一个原子的绝大部分是一无所有的空间，加上量子因素排除了电子具有精确的轨道的可能性，原子便让人觉得像是一种非物质的模糊的实体了。

电磁力使电子离不开原子核，原子核带有正电荷，原子核的周围是电场，电场使带负电荷的电子逃脱不了。很久以前人们就发现，原子核本身也是复合体，是由两种粒子组成的，一种粒子是质子，带正电荷，另一种粒子是中子，既不带正电荷，也不带负电荷。质子和中子的质量分别都是电子质量的1800倍。

物理学家们一旦发现了原子的基本构造，就能够将量子论应用于原子，从而揭示出一种惊人的和谐。电子的波的性质通过电子存在于其中的某种固定的“定态”或“能级”中并将它自己表现出来。假如能量以光子的形式（小批的光能）被吸收或被发射，就会发生能级之间的跃迁。因此，能级的存在是以光能的形式显示出来的，而光能的情况可以从光的频率（颜色）推断出来。分析一下原子所吸收或发射的光，就会发现一种光谱，光谱是由一系列谱线或不连续的频率组成的。最简单的原子是氢原子，是由一个质子（原子核）和一个电子组成的。氢原子的能级可用一个简单的公式

$$\frac{1}{n^2}-\frac{1}{m^2}$$

乘以一个固定的能量单位表示。式中的n和m是整数1，2，3，……。这种紧凑简洁的算术式使人想起音乐的音调，比如吉他或管风琴上的泛音，这些音也可以用简单的数字关系来表示。这并不是偶然的巧合。一个原子的能级的排列是与量子波振动相对应的，正如一部乐器的频率是与声音的振动相对应一样。

假如在氢原子中使电子束缚于质子的力在数学上不是简单的，那么，原子的和谐就不会如此完美。实际上，原子就是依存于这种和谐的。使电子束缚于原子核的电引力满足一个叫作平方反比定律的有名的物理学定律。这个定律说的是，假如质子和电子之间的距离加倍，二者之间的电引力就降低为原来的四分之一；假如二者之间的距离是原先的3倍，二者之间的电引力就降低到原先的九分之一，依此类推。这种井然有序的数学规律也可以在引力中见到，例如，行星和太阳之间的引力就是这样。平方反比定律导致了太阳系的引人注目的规律性，这些规律可用算式表示。运用这些算式，就可以预测日食月食以及其他的天文现象。在原子中，这些规律是量子性质的，表现为能级的排列和发射的光的频谱。但太阳系的规律性和原子的规律性都来自平方反比定律的质朴性。

明白了原子核的结构之后，物理学家们接着就开始探寻原子核内部使原子核结为一体的力。这种力不可能是引力，因为引力太弱，也不可能是电磁力，因为同性的电荷是相斥的，所以，带有正电荷的质子如何竟能在一起相安无事就成了一个谜。显然，必定有一种很强的吸引力来克服质子之间的斥力。实验表明，使原子核成为一体的力要比电磁力强得多，这种力在质子的一定距离或范围之外就突然消失

了。这距离很短，比原子核还要小，所以，只有最近的粒子才处于核作用力的范围之内。中子和质子都处于核力的影响之下。因为这种力很强，所以大多数原子核要用很大的力量才能破开，但要想破开还是办得到的。重原子核不那么稳定，可以很容易地裂变，放出能量。

核粒子也是按不连续的量子能级排列的，只是没有原子和谐的那种质朴性。原子核是一种复杂的结构，这不仅是因为组成原子核的粒子数目多，而且也是因为核作用力并不遵守质朴的平方反比定律。

20世纪30年代，物理学家们在量子论的框架中研究核作用力，终于明白了这种力的性质是与粒子的结构分不开的。在日常生活中，我们把物质和力看成是两个独立的概念。力可以通过引力或电磁效应作用于两个物体之间，或直接通过接触作用于物体。但物质只是被看作是力的来源，而不是力的传播媒介。因此，太阳引力跨越一无所有的空间作用于地球，用场的语言来描述就是：太阳的引力场（若是没有引力表现出来，引力场是看不见也摸不着的）与地球相互作用，对地球施加了一种力。

在亚核的领域里，量子效应起着重要的作用，因而就需要另一套描述及语言。量子论的一个中心论点是，能量是以不连续的量的方式传导的。这也是量子论的由来。因此，光子就是电磁场的量子。当两个带电粒子互相靠近时，就受到了它们都有的电磁场的影响，电磁力就在它们中间起作用。电磁场使它们的运动轨迹发生偏转。但一个粒子通过场对另一个粒子所施加的扰动必须以光子的形式传导。因而，带电粒子之间的相互作用不是一个连续的过程，而应被看作是由一个或多个光子转移造成的脉冲。

在这里，使用理查德·费恩曼所发明的图解有助于说明问题。图23上有一个光子往来于两个电子之间，因而这两个电子便分离开来。有人把这种相互作用的机制比作两个打网球的人，这两个人的行为通过球的往来而有了联系。因此，光子的行为颇像是信使，在两个带电粒

子之间来回跳荡，告诉这个带电粒子那里还有一个带电粒子，从而引起一种反应。物理学家们借助于这样的概念，就可以计算出原子层面上的很多电磁过程的效应。在所有的场合中，实验的结果与利用计算得来的预测惊人地相符。

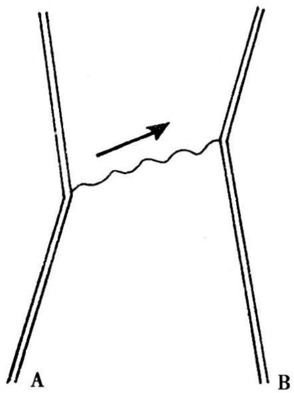



图 23　在量子层面上，带电粒子 A 与 B 之间的电磁力被认为是光子的交换或转移。当光子发射时，粒子 A 的运动路线便显示出一种回弹。后来，B 因为吸收了光子而偏转。粒子之间的力就这样借助其他的粒子（在这里是光子）传导。实际上，这里的描述是有些简化了的。力的传导涉及许多复杂的粒子作用。这些粒子是短寿的（或说是“虚的”），双向运动的，而且各自还绕着 A 和 B 运动。大自然的其他基本的力也可以如此描述。这一类的图象征性地表示了那些抽象的数学式，这些数学式可用来精确地计算亚原子的作用。

电磁场的量子论应用起来如此成功，于是20世纪30年代的物理学家们很自然地又把它应用于核力场。日本物理学家汤川秀树应用量子论，发现质子和中子之间的力实际上可以用二者之间信使般往来的量子为模型，但这里的量子与我们所熟悉的光子大不相同。汤川的量子必须有质量，才能再现出核力的那种作用距离极短的效果。

这里有一个微妙而又重要的问题。一个粒子的质量，就是其惯性

的大小，也就是保持其运动状态不变的力的大小。施加一个同样的力，一个轻粒子要比一个重粒子容易推动。假如一个粒子变得极轻，那它就会被任何杂散的力加速，于是就会以非常大的速度运动。在那极端的场合下，粒子的质量降低为零，粒子就会以最快的速度运行，这速度就是光速。光子就是这种情况，因而可以认为光子是没有质量的粒子。而汤川的粒子则有质量，其运行速度比光慢。汤川把它们称作介子，但现在人们把它们叫作π介子。

π介子在原子核里，往来于中子和质子之间，用核力使中子、质子粘结在一起。通常，π介子是看不见的，因为它们一产生，就接着被另一个核粒子吸收了。然而，假如向原子核系统中输入能量，π介子就能从原子核中飞出来，使人能够单独地对它进行研究。两个质子高速相撞时（这个过程在第三章曾简略地讲过）π介子就会飞出来。第二次世界大战结束不久，π介子就这么被发现了。π介子的发现，出色地验证了汤川的理论，并被誉为理论物理学、尤其是量子场论的胜利。π介子的另一个与众不同的特色是极不稳定，在产生之后几乎立刻衰变成为较轻的粒子。其衰变而成的粒子之一是μ子，这种粒子在各方面都与电子相同，只是质量与电子不同。μ子要比电子重许多，而且也很快就衰变。

物理学家们一旦意识到，通过亚原子粒子的高速碰撞可以造出全新的物质裂片，他们就开始建造巨大的加速器来制造物质的裂变。这些加速器可以把任何一种亚原子粒子加速到接近光速，而接近光速的冲击为人们揭示了亚核行为的整个新世界。这些加速器一旦投入使用，便出现了几十个迄今为止人们未曾想到的新粒子。这些新粒子蜂拥而至，使物理学家很快连名称都来不及给它们取了。一时间，各种各样的粒子乱哄哄地像个乱了套的动物园。后来，物理学家们渐渐地不那么晕头转向了，于是便在亚原子碎片中看出了某种秩序。图样开始出现了。

自20世纪30年代以来人们便知道，核力不是一种，而是两种。强

力将核粒子粘在一起，但还有一种弱得多的力。弱力使某些不稳定的核粒子衰变，例如，π介子和μ子就是由于弱力而衰变的，有些粒子既能感受到强力，又能感受到弱力，但有些粒子则感受不到强力。这种感受不到强力的粒子一般都比较轻，包括μ子、电子和中微子。至少存在着两种中微子，它们都是让科学最捉摸不透的东西。它们与其他物质的相互作用是如此之弱，以至可以轻而易举地穿透好几光年厚的固体铅！

这些相互作用弱的较轻粒子都被称作轻子。带电荷的轻子如电子，既能感受到弱力也能感受到电磁力。但不带电荷的中微子则不受电磁力的影响。相互作用强的较重粒子被称作强子。强子分两种：一种是质子和中子，以及许多衰变为质子和中子的较重粒子，这一种强子被称作重子；其余的强子是介子，包括π介子。

在这些大致的粒子分类中，还可以发现很多亚类。组成这些亚类的粒子具有若干性质，如质量，电荷，以及其他一些更为技术性的特性，其性质随其种类的不同而呈现出有系统的变化。在20世纪60年代，理论物理学家们发现，这些成系统的性质可以用一个数学的分支——群论——来给以非常漂亮的表达。其中的原理是对称的概念；或许可以这么说，物理学界一旦最终意识到了亚原子粒子的对称性，于是便勇往直前了。

人们一直就知道，对称在组织自然界的过程中扮演了一个很重要的角色。我们都熟悉太阳的圆形，雪花和结晶体的规则性。然而，并非所有的对称都是几何性的。男女的对称、正负电荷的对称也是很有用的概念，但这种对称是抽象的。在重子和介子当中也发现了这种抽象的对称，这表明任何特定的一类粒子都被一个简单的数学图表紧密地联系起来。可以用我们所熟悉的几何对称来对此做些许说明。我们都知道，从镜子里看，我们的左手是在右边。左手和右手构成了一个由两个组元组成的对称系统，而镜中的左右手映像又使我们看到了原来的手的样子。从某种意义上说，质子和中子也可以被看成类似左手

和右手。在“映像”中，中子变成了质子，质子变成了中子。当然，这里所说的映像不是通常意义上的在实在的空间里的映像，而是在想象的空间里的一种抽象的映像。这想象的空间用行话说就是同位旋空间。尽管这对称是抽象的，然而，其数学表达却与几何对称是一样的，而且这表达具有足够的真实。在散射实验中的质子和中子的性质，以及质子和中子吸引其他粒子注意的方式，就显示出这种表达是真实的。

更为复杂的对称群，使人们得以对粒子的一些大家族而不仅仅是质子和中子进行统一的描述。某些粒子家族包含有8个、10个或更多的粒子。某些对称乍看之下不明显，因为这些对称被复杂的作用掩盖起来了。但通过数学分析和仔细的实验就可以把它们揭示出来。

这些抽象的对称所显露的物质内部构造的优美，使大部分物理学家感到惊奇。对亚核粒子进行研究的全部基础就是一种坚定的信仰：质朴性存在于一切自然的复杂性之中。尤瓦尔·尼曼和莫里·盖尔曼最先发现，在一个由8个介子构成的集合中隐藏着对称性。他们于是仿照佛陀的话，把他们的新原理称作“八正道”：“这雅利安八正道就是正见，正志，正语，正业，正命，正精进，正念，正定。”

随着越来越多的对称被揭示出来，粒子物理学家们被其精微的规则性深深地吸引住了。这些规则性自天地开创以来就掩藏在原子的深处，不为外界所知。现在，人类才第一次借助先进技术的令人眼花缭乱的器具看到了这些规则性。

物理学家们不久便开始发问这些对称性背后的意义。一位杰出的理论物理学家说：“大自然似乎是想用这些对称来告诉我们什么秘密。”数学分析的力量在这个时候显露了出来。群论表明，一切对称都可以在一个单一的主要基本对称中找到其自然的起源。人们发现，较为复杂的对称都可以通过非常简单的组合得到。用粒子研究的术语来说，数学表明，强子根本就不是基本粒子，而是由更小的粒子组成

的合成物。

这的确是轮中之轮！原子是由原子核和电子构成的，原子核是由质子和中子构成的，质子和中子是由什么构成的呢？这些新发现的物质的基本构件构成了质子和中子，它们与原子隔了三层，当时还没有名称。盖尔曼于是便杜撰了一个名称——夸克。而这名称还真的就这么叫上了。强子是由夸克构成的。古希腊人认为，一切物质都是由为数不多的基本粒子（即他们所谓的“原子”）构成的。这一伟大的原理已被事实证明不那么好理解。基本粒子是否就是夸克？难道夸克也是复合体吗？我们一会儿再来讨论这个问题。

夸克以两种构型附着在一起。一种构型是两个夸克附在一起，另一种构型是三个夸克附在一起。两个夸克在一起就构成了介子，三个夸克在一起就构成了重子。夸克也有量子能级。能够通过吸收能量而受激进入较高的级位。受激的强子看上去与其他的强子一样，于是，很多先前被认为是独立的粒子现在被看作是单个夸克结合的受激状态。

为了解释所有已知的强子，就必须设想夸克不止一种。在20世纪70年代初，人们设想有三种“味道”的夸克。这三种夸克被异想天开地称作“上”、“下”、“奇”。后来，出现了更多的强子，又多出了第四种夸克，即“粲”夸克。近来，出现了更多的粒子，人们认为还得有另外两种夸克：“顶”夸克和“底”夸克。现在，很多种粒子作用都可以借助详细的夸克计算获得系统的了解。

夸克理论的基本预设是，夸克本身是真正浑然一体的基本粒子，是一种像点一样的物体，没有内部成分。在这方面，夸克颇像轻子，因为轻子不是由夸克组成的，它们本身似乎就是基本粒子。事实上，夸克和轻子之间有着自然的对应，使人们获得意想不到的机会得以洞见大自然的运作。夸克和轻子之间的系统联系见下面的表1。表右边一栏是夸克的味道，左边是已知的所有轻子。要记住，轻子感受到的

是弱力，而夸克感受到的是强力。轻子和夸克之间还有一个区别是，轻子或是不带电，或是只带1个单位的电荷；而夸克则带三分之一或三分之二单位的电荷。

**表 1**

| | 轻子 | | 夸克 | |
|---|---|---|---|---|
| | 名称 | 电荷 | 名称 | 电荷 |
| Ⅰ | 电子（e） | −1 | 上（u） | $+\frac{2}{3}$ |
| | 电中微子（Ve） | 0 | 下（d） | $-\frac{1}{3}$ |
| Ⅱ | $\mu$子（$\mu$） | −1 | 奇（S） | $-\frac{1}{3}$ |
| | $\mu$中微子（$\upsilon_{\mu}$） | 0 | 粲（C） | $+\frac{2}{3}$ |
| Ⅲ | $\tau$子（T） | −1 | 顶（t） | $+\frac{2}{3}$ |
| | T中微子（$\upsilon_{T}$） | 0 | 底（b） | $-\frac{2}{3}$ |
| | ? | ? | ? | ? |

亚原子粒子可分为两大类：轻子和夸克。夸克没有被发现单独存在，而是两个或三个地在一起。夸克的电荷是分数的。一切普通的物质都是由Ⅰ层面的粒子构成的。Ⅱ层面和Ⅲ层面似乎是Ⅰ层面的简单复制，其中的粒子是高度不稳定的。可能尚有未发现的层面。

表中没有列入信使粒子：光子、引力子、胶子以及弱作用力的介体W和Z。

尽管轻子与夸克有着如此的差别，但二者之间存在着深刻的数学对称，使轻子和夸克在上面的图表中有了逐层面的对应。第一个层面只有4种粒子：上、下夸克、电子及电中微子。奇怪的是，一切普通的物质竟全是这四种粒子构成的。质子和中子是由三个三个的夸克组成的，而电子只是充任构成物质的一种亚原子粒子。中微子只是跑进宇宙里，一点也不参与物质的大体构造。就我们所知，假如其他的粒子都突然消失了，只要有这4种粒子，宇宙就不会有多大变化。

下面一个层面的粒子似乎就是第一个层面的复制，只不过较重而

已。第二个层面的粒子都极不稳定（中微子例外），它们所构成的各种粒子很快就衰变为层面I的粒子。第三个层面的粒子也是这样。

于是就必然产生这样的问题：层面I之外的其他粒子有什么用处呢？为什么大自然需要它们？在形成宇宙的过程中，它们扮演了什么角色？它们是多余的赘物？或者，它们是某种神秘的、现在尚未完全明了的过程的一部分？更为令人不解的问题是，随着将来能量越来越高的粒子加速器的出现，是否也只有这三个层面的粒子？是否会发现更多的或无穷多的层面？

还有一种复杂的情况加深了我们的不解。为了避免与量子物理学的一个基本原则相冲突，我们必须设想每一种味道的夸克实际上有三种不同的形式，即人们所说的“颜色”。任何一个给定的夸克都必须被看作某种多层电镀（比喻说法）的叠加，不断地闪现出（又是一个比喻说法）“红”、“绿”、“蓝”的颜色。这样一来，一切又看上去像是乱了套的动物园了。但是，收拾局面的方法就在眼前。对称又来救驾了。不过，这一次的对称，其形式更微妙、更深奥，怪不得被人们称作超对称。

为了理解超对称，我们就得说说物质基本结构分析的另一个大线索：力。不管粒子动物园有多么纷纭复杂，其中看来只有4种基本的力：引力，电磁力（因与日常生活密切相关而广为人知），弱作用力和强作用力。中子和质子之间的强力，当然不可能是基本力，因为中子和质子本身就是复合物而不是基本粒子。当两个质子相互吸引时，我们实际上看到的，就是6种夸克相互作用的合力。夸克之间的力才是基本力。可以用描述电磁场的方式描述夸克之间的力，而夸克的颜色就相当于电荷。质子的对应物是所谓的“胶子”，其作用就是我们先前说过的像信使那样，不断地在夸克之间来回跳动，将夸克胶结在一起。物理学家们仿照电动力学，把这种由“颜色”产生出来的力场理论叫作色动力学。色动力作用要比电磁力作用复杂。这有两个原因。第一，夸克有三色，而电荷却只有一种，于是，与一种光子相对应的就

是8种不同的胶子。第二，胶子也有颜色，因而彼此也有很强的相互作用，而光子不带电荷，彼此间又是那么不相干。

20多年前，某些富有远见的理论物理学家突然想到，大自然有四种基本力，这数目似乎太多了。很可能这四种基本力并不是真正独立的。麦克斯韦在19世纪60年代提出了一个数学式，使电力和磁力统一于一个单一的电磁场理论。很可能还会有进一步的综合。

一种徘徊不去的难以解决的数学问题更推动了某些理论物理学家做如是想。除了最简单的作用之外，每当人们把量子论应用于所有的作用时，得到的结果总是无穷，因而也就是无意义的。而将量子论应用于电磁场时，有一种数学特技使人们能够绕开无穷，量子论因而也就一直能预测一切人们所能想象的电磁作用。但同一个数学特技对其他三种力却不灵。人们希望，通过某种方式把电磁力和其他三种基本力结合进一个单一的描述式，这一个单一的描述式所具有的数学温顺性会消解电磁力之外的其他三种力，使人们能够得出一个可以理解的算式。

实现这一宏伟目标的第一步是斯蒂芬·温伯格和阿布杜斯·萨拉姆在1967年迈出的。他们成功地改造了电磁力和弱作用力的数学表达式，使这两种力被结合进一个统一的数学表达式之中。他们的理论表明，我们通常之所以把电磁力和弱作用力看成是不同的力（确实，二者在性质上显著不同），是因为在我们现行的实验中所利用的能量极低。当然，这里所说的“低”是相对而言：现在的加速器可以给一次对撞足够大的能量，假如这能量不是加在一个质子上而是加在一个台球上的话，释放出来的能量就能为一个普通人家提供几百万年之需！不过，温伯格一萨拉姆理论有一种内含的能量单位，这种单位的能量只是到了现在才能由现有的技术达到。上面所说的现行实验所利用的能量“低”，也是与这种单位相对而言的。

在20世纪70年代，实验的证据慢慢积累起来，情况变得有利于温

伯格一萨拉姆理论。1980年，他们因为统一力研究方面的工作获得了诺贝尔奖。1971年就已经证明，那令人头痛的无穷可以像所希望的那样，在一个统一式中被扫除，物理学家们开始谈论大自然的三种而不是四种基本力了。

那令人头痛的无穷之所以能被扫除，其主要原因是在统一力的理论中出现了更加抽象的对称群。人们早就知道，麦克斯韦优美的电磁理论之所以有力量，之所以优美，在很大程度上要归功于该理论的数学描述中所显示出来的平衡和对称。统一力的理论中又来了平衡，这平衡被称作规范对称，是一种抽象的平衡。但这种平衡能让人想起日常生活中的事。

可以用攀登断崖的例子来说明规范对称。从崖底攀到崖顶要耗费能量。但是，由下往上攀登有两条途径。一条较短，是垂直着直接登上崖顶；再一条较长，是顺着较缓的坡道登上崖顶。这两条途径哪一条更有效率呢？（见图24）回答是：两条途径都要耗费相同的能量（在这里，我们对诸如摩擦之类不相关的复杂情况忽略未计）。实际上很容易证明，攀登崖顶所需的能量是与所选用的途径完全无关的。这就是规范对称。

上面所举的例子说的是引力场的一个规范对称，因为你要攀上崖顶，必须克服的是引力。规范对称适用于电场，也适用于与电场类似但更为复杂的磁场。

现已证明，电磁场的规范对称是与光子没有质量的特性密切相关的，同时，也是使统一力理论避开灾难性的无穷的一个关键性因素。温伯格和萨拉姆终于驯服了弱力，使之与电磁力合并起来。

物理学家们受到统一规范理论成功的鼓舞，把注意力转向了另一种核力——夸克间的色动力。不久之后，就提出了色规范理论，接着，有人便试图将弱力和色动力统一到一个“大统一理论

（（GUT）”中去，办法是使用更大的规范对称将所有的其他对称包容在一个规范对称之中。目前，估价GUT的成就还为时尚早，但至少它所作的一个预测经过无限长的时间之后，质子可能会很不稳定并自发地衰变一一现在正有人进行检验。

但是，引力仍是没有就范。无穷的难题报复性地缠住引力不放。现在，物理学家越来越倾向认为，只有在包含了某种超对称的一种超统一理论中，这一难题才会获得解决。一大群数学家和物理学家正在为创制一个这样的理论而奔忙。这一理论的目标，是那不可抗拒的统一场理论的梦想——一个单一的力场，涵盖大自然的所有的力：引力，电磁力，弱作用力和强作用力。但是，这还远远不够。量子粒子和作用于这些粒子之间的力表明，任何一种力的理论同时也是一个粒子的理论。那么，超统一理论也应当能完全描述一切夸克和轻子，解释为什么在表1中有三个层面的粒子。

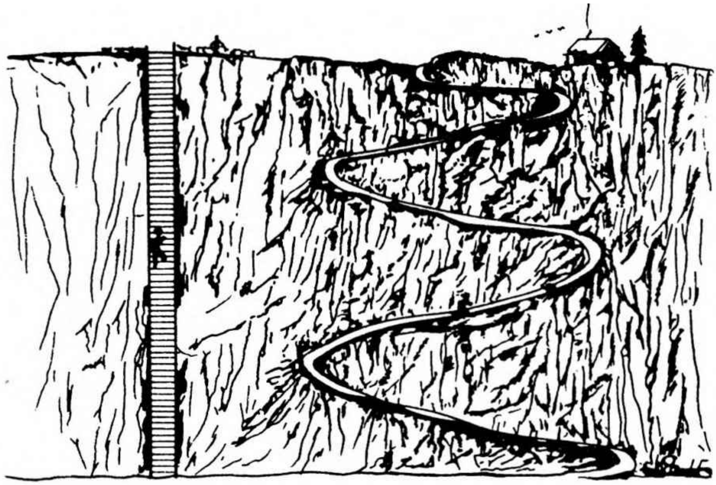


图 24　攀登断崖可以用来说明“规范对称”这一抽象的概念。不管是走那条直而短然而却是艰难的路，还是走那条长而易走的之字形路，攀上崖顶所需的全部能量是相同的。这反映的就是引力场深刻而有力的对称。大自然其他的力场与此相似而更为复杂的对称，最近在统一场理论的数学表达中得到了利用。

有人说，要是真能达到这个令人目眩的目标，也就是达到了基本物理学的顶点，因为像超统一理论这样的一个理论能够解释一切物质的行为和结构——当然，是以一种还原论的方式进行解释。有了超统一理论，我们就能够用一个方程式，用一种宇宙的总公式把大自然的一切秘密都写下来。这样的一个成就会证实人们长久以来所宠爱的信仰——宇宙是按照一个单一的、质朴的、具有惊人优美的数学原理运行的。约翰·惠勒下面的话，就表达了人们要达到这一最终目标的迫切心情：“总有一天，有一扇门肯定会开启，显露出这个世界的闪闪发光的中心机制，既质朴，又优美。”①

我们离这智慧的极乐世界还有多远呢？理论物理学家们现在正把他们的希望押在一套理论上。这套理论的名称叫超引力。这套理论的关键是一种奇异的超对称，这超对称被描述为时空的平方根。它的意

思是，假如两个超对称运算式相乘，你就会得到一个普通的几何对称运算，如空间中的移动。

乍看之下，这种抽象似乎没有什么大用处，但仔细分析就可以看到，超对称与一个粒子可能具有的最基本的属性之一——旋转——有着密切的关系。人们发现，所有的夸克和轻子都以一种颇为神秘的方式旋转。我们现在且不去管它如何旋转。我们要关心的是，那些“信使”粒子——胶子、光子、还有引力和弱力的相应的粒子——或者是不旋转，或者是以一种正常的而不是神秘的方式旋转。超对称的意义就在于，它把以神秘的方式进行旋转的粒子和其他的粒子联系了起来，正如同位旋对称把质子和中子联系起来一样。于是，超对称的运作能把一个旋转的粒子变成一个不旋转的粒子。当然，这里所说的“运作”指的是数学步骤。实际上，把一个旋转的粒子变成一个不旋转的粒子是不可能的，正如你不能把你的左手变成右手一样。

通过把引力理论置于超对称的构架之中，引力的信使粒子（称作引力子）就获得了以一种“好玩的”方式旋转的同伴粒子（称作gravitinos），以及其他的粒子。这么多种类的粒子进入超引力理论，这就有力地表明，那可怕的无穷难题被压下去了，而且，到目前为止利用这一理论进行的一切具体运算得出的结果都是有穷的。

在最为人们看好的一种超引力理论中，整个的粒子大家族的成员总数不超过70。这种理论所包含的许多粒子都能够被认定就是现实世界中已知的粒子。不能被认定的粒子则是可能存在但现在尚未发现的粒子。这一理论是否将迄今为止被认作基本粒子的一切粒子都包括进去了？是否实际上可能会有更多的基本粒子？对此，人们的意见尚不统一。有的理论物理学家认为，夸克的数目太多，现在是进一步深入研究，搞明白这么多的夸克是否是由更小的物质单位构成的时候了。对此看法，有人提出了反对意见，认为物质的结构没有比夸克更低的层面了，夸克的世界已经是原子核的大约$\frac{1}{10^{15}}$了，而这离空间失去意

义的那个最终尺度已相去不远了。理论研究显示，在一个原子的约$\frac{1}{10^{20}}$的尺度内，引力的量子效应使时空变成了模糊不清的东西，在这一尺度上谈论什么东西存在于什么东西“之内”就变得无意义了。因此，关于是否还有更基本的物质单位的研究工作仍在进行。

我希望，我对物理学家们正在进行的揭示物质终极结构的工作所做的简略介绍，至少能让大家多少对现代物理学研究有点认识。物理学家对待其研究对象的态度近乎敬畏，因为他们总是受一种信仰的支配，这就是，大自然是由数学的优美和质朴统治的；通过深入探究物质的结构，大自然的统一性将会显明出来。迄今为止的一切经验表明，所探寻的系统越小，所发现的原理就越一般。按照这一经验来看，被我们偶然发现的世界的复杂性，在很大程度上纯是我们的物质取样系统的能量相对较低的结果。人们相信，随着取样系统的能量越来越高，大自然的统一性和质朴性也会变得越来越显明。这也就是为什么这么多的人力物力被投入建造超高能粒子加速器的缘故。人们想通过超高能粒子加速器闯进那质朴的状态去探寻究竟。

然而，曾经有过那么一个时期，当时，这种质朴的状态被大自然探寻过。那时，宇宙在大爆炸中诞生还没有1秒，当时的温度高达$10^{27}$度，正好可以用作探寻原初质朴状态所需的能量。这一段时间，物理学家们称之为大统一时代，因为当时的物理正是受基本力的大统一理论的过程支配的。我们在第三章里所提到的至关重要的非平衡就是在当时确立的，而有了那种非平衡，才导致了物质稍稍多于反物质。后来，随着宇宙的冷却，原初的统一力也分化为三种不同的力——电磁力，弱作用力，强作用力。这些力都是我们在相对冷却下来的宇宙中所看到的。

今天的复杂的物理，是由原初大爆炸火焰构成的质朴的物理冷却而成的。这种看法，倒是美妙而吸引人。大自然的最终原理，也就是惠勒所孜孜以求的“闪光的中心机制”，我们因能量不足而难以窥见。

假如人们追踪到大统一时代以前的那些时期，追到离时间起始处更近、温度更高的地方，就可以找到超引力了。超引力所代表的，就是存在的起始，在起始之处，时间和空间同基本力都结为一体。大多数物理学家认为，时空的概念在超引力时代之内是不能用的。实际上，有迹象显示，时间和空间也应被看作是两种场，这两种场本身也是几何元素组成的原初汤“冷却”而成的。因而，在这超引力的时代中，大自然的四种力是混沌一体的，而时空则尚未成一个像样子的形。当时的宇宙只是一堆超质朴的元件，是一些上帝用以造出时间、空间和物质的原料。

本章描述了物理学关于基本力研究的新近进展。这些进展已使人们以全新的观点看待大自然。这种观点的影响在物理学家和天文学家中间迅速扩大。现在，人们已开始把宇宙看成是由质朴的东西冷却而生成的复杂的东西，颇像是浑然无形的海洋冻成了姿态各异的浮冰。科学家们有一种感觉，这就是宇宙学的研究课题和人们对物质当中的基本力的研究正在为宇宙提供一个统一的描述。在这种描述中，物质的极微结构与宇宙的总体结构紧密联系在一起，两种结构都以一种微妙而复杂的方式影响着彼此的发展。

本章所描述的物理学的一系列成功，无疑代表了以还原论理论为其基础的现代物理学思想的一个胜利。物理学家们试图把物质还原为最终的构件——轻子、夸克、信使粒子一一从而得以瞥见那基本的定律。而正是那基本的定律控制着形成物质的结构和行为的力量，从而能够解释宇宙的很多基本特点。

尽管如此，以这种方式追寻某种已被感觉到的终极真理是远远不够的。我们在前面的几章里看到，还原论不能够解释很多明显的具有整体性特征的现象。例如，我们不能用夸克来理解意识、活的细胞，甚至也不能以之理解诸如龙卷风之类的无生命的系统。否则，一定会闹出笑话的。

到目前为止，本章所用的语言在很大程度上并没有传达出物理学家心目中的物质结构的概念。当一个物理学家说，质子是由夸克“组成的”时，他的本意并非如此。比如，我们说一个动物是由细胞组成的，或一个图书馆是由书组成时，我们的意思是说我们可以拿来一个细胞或一本书，或从那较大的系统那里随便拿来什么东西，进行孤立的研究。但夸克却不是这样。就我们所知，不可能真的拆开质子拿出夸克来。

然而，拆开有着辉煌的历史。拆开原子现在已成了家常便饭；原子核敲开较难，但在高能的冲击下也会分裂。这或许意味着用高速粒子轰击质子或中子，将会把质子或中子粉碎为夸克。然而，实际情况却不是这么回事。一个极小的高速电子会穿过质子的内部，将其中的一个夸克猛烈地弹开，从而使我们确信质子内部的什么地方确有夸克。但是，若打击质子的不是小小的电子，而是一个大锤，即另一个质子，那么，我们就不会在质子的碎片中看见夸克，而只能看见更多的强子（质子、介子等）。换言之，夸克从不孤立地出现。大自然似乎只准许夸克以集体的面目出现，出现的时候总是两个两个或三个三个地在一起。

因此，当物理学家说质子是由夸克组成的时，他的意思并不是说这些神秘的夸克可以单独地显现出来。他只是指一个描述层面，这一层面比质子层面更基本。管辖夸克的数学法则要比管辖质子的更质朴、更基本。从某种意义上说，质子是合成的，不是基本的；但质子由夸克的合成与图书馆由图书的合成不是一码事。

当我们像在第八章里看到的那样，将量子因素纳入考虑之中时，理解物质的基本结构便有了更为严重的困难。这是因为，没有哪种亚原子粒子（不管是夸克还是什么别的基本粒子）是货真价实的粒子。实际上，亚原子粒子可能连“东西”都算不上。这就使我们又一次认识到，所谓物质是某某粒子的集合这种描述，实际上必须被看作是由数学所确定的描述层次。物理学家对物质结构的精确描述只能通过抽象

的高等数学来进行，而人们只有认识到这一背景，才能明白还原论所说的“由……组成”的真正含义。

海森伯的不确定性原理的一个方面，很好地说明了量子因素给研究“什么是由什么组成的”这一课题带来的困难。但这次的二象性，不是波粒之间的二象性，也不是运动与位置的二象性，而是能量与时间之间的二象性。能量与时间这两个概念处于一种神秘莫测的对立关系之中：你知道了一个就不知道另一个。因而，哪怕在一个很短的时间内观察一个系统，其能量也有可能发生巨大的起伏。在日常的世界里，能量总是守恒的。能量守恒是经典物理学的柱石。但在量子微观世界里，能量可能以自发的、不可预测的方式不知从哪里冒出来，或消失在哪里。

当考虑到爱因斯坦著名的E＝mc2的公式时，量子能量的起伏就变成了复杂的结构。爱因斯坦的公式说的是，能量和质量是相等的，或者能量能够创造物质。这已在前几章里讨论过了。不过，那几章里所说的能量来自外部。这里，我们想讨论一下，在没有外部能量输入的情况下，物质粒子如何能从量子能量的起伏中被创造出来。海森伯的原理颇像个能量库。能量可以短期借用，只要迅速归还就行。借用期越短，可借用的量就越大。

比如在微观世界中，一次突然的能量起伏可能使一个正负电子对在短期内出现又消失。这正负电子对的短暂存在，就是由海森伯式的借贷维持的。其存在的时间从不超过$\frac{1}{10^{21}}$秒。但无数个这样忽隐忽现的幽灵粒子累加起来的效果，就使空无一物的空间有了某种变换的质地，尽管这是一种模糊的、不实在的质地。亚原子粒子就必须在这不停运动的海洋中游动。不仅电子和正电子，而且质子和反质子，中子和反中子，介子和反介子，总之，大自然的所有粒子都是这么动荡不安。

从量子的角度来看，一个电子不仅仅是一个电子。变换能量的花

样在其周围闪烁着，不知什么时候突然促成了光子、质子、介子甚至其他电子的出现。总之，亚原子世界的一切都附着在电子上，像是电子穿上了看不见摸不着的、转瞬即逝转瞬又来的一件大衣，或者说，像是幽灵一样的群蜂嗡嗡地围着中间的蜂巢飞翔，构成了蜂巢的覆盖物。当两个电子相互靠近时，它们的覆盖物也纠缠在一起，于是，相互作用就发生了。所谓的覆盖物，只不过是将先前被看作是力场的东西加以量子的表达罢了。

我们永远也不能将电子跟其所带有的幽灵粒子分离开来。当有人问“什么是电子”时，我们不能说电子就是那个小粒子；我们必须说电子是不可分离的一整串东西，包括跟它在一起的产生力的幽灵粒子。说到具有内部结构的强子，就更加模糊难辨了。一个质子不知为何总是带着夸克，而夸克又是由胶子连在一起的。这里也有一种怪圈：力由粒子产生，而被产生的力又产生力……

而对光子这样的粒子来说，这种怪圈意味着光子可以展现出很多不同的面孔（faces）。通过借入能量，它可以暂时变成一个正负电子对，或一个正反质子对。已有人进行了实验，试图看到光子是如何变成正负电子对或正反质子对的。但是，人们又一次发现，要想从这种错综复杂的变化中分离出“纯”光子是不可能的。

就大多数不稳定而且寿命又极短的粒子来说，已难以说清哪些是“实在的”，哪些是“幽灵”。有一种ψ粒子，在$\frac{1}{10^{21}}$秒内就衰变了；而由海森伯原理造成的正负电子对，其寿命也跟ψ粒子差不多。谁能说前者是实在的，后者只是个幽灵呢？

一些年前，一位叫杰弗里·邱的美国物理学家把亚原子世界中的这种闪烁不停的变幻比作一个民主政体。我们不可能抓住一个粒子，说它就是某某实体。我们必须把每一个粒子看成是在一个没有终结的怪圈中由所有的其他粒子组成的。没有哪一个粒子比其他任何粒子更基本（这就是我们在第四章里简短地提到过的“拽靴襻”）。

我们将会看到，物质的本性在其量子论方面具有强烈的整体论的味道：物质的不同层面的描述是相互连锁的，一切东西都是由另外的一切东西组成的，然而一切东西同时又显示出结构的等级次序。物理学家们就是在这无所不包的整体性中追寻物质的终极成分，追寻终极的、统一的力。

# 第十二章　偶然还是设计？

我们在这世界上所看到的这一切秩序，这一切美，又是从哪里来的呢？

艾萨克·牛顿

人类终于知道了，他们在这广漠无垠、没有感觉的宇宙中是孤立无援的……他们的命运、他们的义务都没有被明文规定下来。

《偶然与必然》雅克·莫诺

威廉·佩利（1743～1805）在其《自然神学》一书中说的下面的话，是上帝存在的最有力的论证之一：

走过一块荒地时，设若我的脚碰上了一块石头，有人问我那石头是怎么到那里去的，我很可能回答说，那石头一直就在那里，因为我不知道它曾不在那里。要证明这样的回答是荒谬的，可不是件很容易的事。但是，设若我在地上发现一只表，有人问我表是怎么到那里去的，我就几乎不会用前面的答话来回答这一次的问题，我不会说，据我所知，表可能一直就在那里。可是，为什么仅因为一个是石头，一个是表，回答就该是两样呢？①

表的构造精微复杂，各个部件衔接精确。它不容置疑地显示了人的设计。即使一个从未见过表的人见了表之后，也会得出结论说，这种机械装置是一个有智慧的人为了一定的目的而设计出来的。佩利接着论证道，就其构造和复杂性而言，宇宙就像一只表，只不过比表大得多罢了。因而，肯定是有一位宇宙设计者为了某一目的把世界安排

成这样——“大自然的构造，就复杂性、就结构的精巧而言，超过了人工的构造。”

宇宙出自设计这样的论证于是跟目的论挂上了钩。因为目的论就是认为宇宙是按照定好的程序向着某个最终的目的演化的。目的论在其最广泛的形式中包含了质朴的秩序和复杂的秩序。目的论是一种古老的观念。阿奎那曾经写道：“人们在一切物体中都观察到趋向某个目的的行为秩序，一切物体都遵从自然规律，即使当它们没有意识时也是如此……这就表明，它们确实趋向一个目的，而不是偶然地碰上目的。”尽管阿奎那对物理基本定律的数学质朴性一无所知，但他点出了物体遵从秩序的规律这一引入注目的事实，并以之作为设计者上帝存在的证明。

目的论曾受到激烈的攻击，以致现在神学家们也对目的论怀有戒心。然而，一些现代人倒为目的论作辩护。斯温伯恩写道：“宇宙中存在着秩序，这就显然地增加了上帝存在的可能性。”②但是，斯温伯恩立论的基础是质朴的秩序，而不是复杂的秩序。复杂的大自然的结构证明有一个宇宙的设计者存在，这种论点似乎已经声名狼藉了。

这种论点之所以声名狼藉，主要是因为很多显示出复杂秩序的系统实际上可被解释为是由完全普通的自然作用所造成的最终结果。当然，这并不是说，一切有序系统都是自然地产生的，但这也的确使我们小心起来，不能仅仅因为看到某种事物很复杂，不像是偶然产生的，就推论说存在着一个设计者。我们也必须了解一些复杂的秩序得以产生的过程。

随着查尔斯·达尔文《物种起源》一书的出版，目的论与反目的论的重大冲突就产生了。生物的精巧组织似乎最充分地显明了一个超自然的设计者的存在，而生物学以及地质学则为生物的所有的不寻常的特性提供了充足的解释。现在，科学家和神学家实际都一致认为，生物界的秩序的演化，是由突变和自然选择造成的。尽管达尔文最初的

理论到现在也没有完善，但进化的基本原理和机制则没有人去认真地怀疑了。

达尔文的进化论的主要论点是偶然性。突变是由纯粹的偶然造成的，由于生物特性中发生的这些完全随机的变化，大自然就有了广阔的选择范围，可以根据适应性以及优越性进行选择。这样，大量的小偶然变易积累起来，就产生了复杂的有组织的结构。这种趋势所引起的相应的有序的增长（熵的降低）是以更大量的有害突变为代价的。通过自然选择，有害的突变被除去了。因而，生物的进化与热力学第二定律并不矛盾。今天的美妙的生物是靠着遗传灾难作铺垫发展起来的。

不管人们是否准备承认达尔文提出的进化论机制是完善的，但不可否认的是，突变和自然选择肯定是促成生物秩序发展的一个主要因素。物质系统可以自发地组织起来，形成错综的复杂性，这一至关重要的原理是一个经验的事实。在第五章我们曾看到过，近年来物理学家和化学家如何在实验室里研究较简单的自组织的例子。实际上，这些研究变得如此重要，以致人们造出一个新词——协同学——来描述这些研究。所得出的结论必然是，一个系统当中所存在的秩序，不管多么引人注目，多么复杂，其本身并不能证明它必定是一个设计者创造出来的。秩序可以而且也确实自发地产生。

然而，这些意见仍没有解决一个重要的问题。尽管只要在其他地方产生代偿性的无序，秩序的自发产生就不会与热力学第二定律相矛盾，然而，假如宇宙作为一个整体在开始时没有相当的负熵储备，显然根本不可能存在任何秩序。假如总体的无序根据热力学第二定律一直是在增加，那么，在我们看来，宇宙创生时必定是有序的。这难道不是为一个创世主——设计者的存在提供了一个强有力的证据吗？因为即使自然的过程可以产生出局部的秩序，但首先仍是需要先有些负熵来驱动这些自然的过程。不错，负熵的存在充其量只能证明有一个代理设计者，即一位创造者给大自然这部机器输满了能量，然后由它

自己随便产生出什么结构来。但是，这样的说法仍是牵涉达到惊人程度的超自然的灵巧。其原因说明如下。

熵，即无序，是与概率和排列的概念密切相关的。一个高熵或无序的系统可能是很多原因的结果。例如，我们可以考虑一下一箱处于平衡状态的气体的情况。箱内的气体现在温度一致，密度一致，达到了最高熵的状态。在这种情况下，气体的所有的分子可以极多的方式重新排列（例如，把分子挪到不同的位置上，或改变它们的运动速度）而不影响气体的大体性质。另一方面，我们再考虑一下低熵状态。我们所考虑的低熵气体分子或以平行的轨道运动，或是都挤在箱子的一边。这些有序的分子排列构型对任何细微的分子重新排列都极其敏感。分子重排列的方式数目极大，但排列出这种有序构型的方式数目却很小。这也就是说，有序（低熵）的状态是高度不可几的，不稳定的。低熵状态要求数目庞大的个体分子进行细致的合作。而处于无序（高熵）状态的分子则可以撇开其他的分子不管，胡乱地运动。

例如现在让你随意挑一种分子排列，那么，极有可能的是，你挑的是具有最高熵的排列。原因很简单，因为可能的无序排列要比有序排列多得多。这颇似一个猴子在乱弹琴，它弹出一首名曲的可能性比弹出不成曲调的一串噪音的可能性要小得多。数学研究表明，有序状态对重新排列的敏感性是呈指数关系的。这就是说，进行一次随机的选择而导致有序状态的几率，随着负熵程序的增长而呈现指数下降。指数关系的特色是其迅速地增长或降低。例如，一些以指数关系增长的生物每隔一段时间数目就会加倍：1，2，4，8，16，32，……

指数因素的存在意味着随机发生有序状态的可能性极小。例如，一个箱子里的1升空气自发地全跑到箱子的一头的几率是$10^{1020}$。这个数字代表的是1后面有100000000000000000000个零！这样的数字说明，从数目庞大的各种可能的状态中挑选出低熵状态（有序状态）必定要多么细心。

这个谜在宇宙学中的意义是这样的：假如宇宙的创生纯属偶然，那么，宇宙中包含任何可观的秩序的可能性便小得不成样子。假如大爆炸只是个随机事件，那么，可能性极大的情况（用“极大”一词极不够分量）似乎就是，随大爆炸产生的宇宙物质将会处于热平衡状态之中，熵值极大，有序程度为零。而事实显然并非如此，于是，人们就很难回避这一结论：宇宙的实际状态是不知用什么方法从数目庞大的可能的状态中“挑选”出来的，因为这些数目庞大的可能的状态除数目极小的一部分之外是完全无序的。假如宇宙这种极不可能的有序的初始状态被选出来了，这岂不就是说当初必定有一个挑选者或设计者进行了“挑选”吗？

这里可以用一个形象进行说明。有个造物主带有一根别针。他前面摆着一大串各种宇宙供他挑选，其中每种宇宙都以其初始状态作其标记。假如这位造物主把别针胡乱别在一个宇宙上，就这样挑出一个宇宙，那么，极有可能的是，他所选择的宇宙是高度无序的，没有可观的结构或组织。事实上，这位造物主若想发现一个有序的宇宙，就必须在一大堆“模型”中进行搜索，而这些“模型”的数目又如此之大，以至在一张大如可见的宇宙的纸上也写不下来。

宇宙是如何进入其低熵状态的？这个谜牵动了好几代物理学家和宇宙学家的想象力，他们当中很多人一直不愿意求助于上帝的选择来解决这一问题。统计热力学的先驱路德维希·玻尔兹曼宁愿认为是盲目的机缘使宇宙进入了低熵的状态。他认为，宇宙的有序状态是由一些对平衡状态的偏离之间的协作造成的。这些偏离十分罕见，罕见得无法想象。他立论的基础是这一事实：即使是在平衡的状态中，气体分子也不是安然不动的，而是不停地以一种随机的方式四处冲撞。可以时时发现，一些分子由于纯粹的巧合而处于无意的合作状态之中，在一个极短的时间里，混沌的大洋里会出现一小块有序的飞地。加倍放大时间尺度，人们便可以相信更大的协作区域将会偶然地最终出现。假如给宇宙足够的时间，那么，人们就可以设想迟早会偶然地形成所

有的恒星，所有的星系。出现这种不可能得近乎荒唐的事件所需的时间长得不可想象（至少得要$10^{1010}$年），不过这没有什么要紧的，假如人们愿意相信宇宙的年龄无限的话。

照这种观点来看，宇宙在全然混沌没有任何组织的状态之中度过了其绝大部分时间。但是经过长得说不上来的间隔之后，宇宙间会出现几十上百亿年的偶然的秩序。我们人类之所以能亲眼看到这种极其不可能的事，只是因为若没有这样的“奇迹”，生命就不可能存在。因为生命是以负熵为生的（见第五章），有意识的观察者就只能存在于宇宙发生“奇迹”、偏离平衡状态的时期。

玻尔兹曼的推理有一个有趣的副产品，这就是它断言存在着某种形式的永恒。可以从数学上证明，使宇宙充满能量的分子的无间歇的往复运动具有下面奇特的特征。随着分子四处乱撞，宇宙也进入一个又一个的状态。最后，所有可能存在的状态都会被宇宙进入一遍，就是说，任何可能会发生的事迟早都会发生。然后，宇宙间的物质继续排列组合，宇宙就会开始重新进入先前有过的状态。最后，所有的状态都会被重新进入一遍，于是，这样的过程就这样持续不已。这种无限重复和复制的现象被称作庞卡莱循环，因为是庞卡莱这位数学物理学家证明了这个结果（至少，他证明一个理想的模型会有这种结果）。假如从字面上看，庞卡莱定理便意味着，在无限充足的时间里，行星地球消失之后，还会重新组合起来，并且连带着住在地球上的居民！而且，这样的事会发生无限次。但是，这种大致精确的复制每发生一次，就会有无数次偏离目前的排列的情况。复制得越是精确，几率也就越小，等的时间也就越长。

玻尔兹曼对宇宙成因的解释，没有几个物理学家愿意相信。庞卡莱所证明的循环的基本机制虽没有受到怀疑，但人们现在知道，宇宙并不是在那里混日子，任其物质随机组合排列。宇宙现在处于一种全面的膨胀状态。人们普遍认为，宇宙的这种全面的膨胀迫使宇宙具有有限的年龄。宇宙区区几百亿年的年龄，比起能够产生一点点熵值降

低所需的时间，完全是沧海一粟，不值一提。

不过，玻尔兹曼的观点确实提出了一个具有永久价值的重要问题。我们所感知的宇宙必然是由我们选择的。因为生命以及由生命而来的意识起码要在合适的物理条件下才能发展起来。明确地说，我们不可能观察一个没有人居住的宇宙。我们马上就会看到，有些人一直利用这一简单的事实来说明，我们所观察的极不可能的低熵宇宙是从众多可能的宇宙中选择出来的（几乎所有的可能的宇宙都是无序的）；但进行选择的是我们，不是上帝。

因而，假如承认有过大爆炸，那么，我们看来就只能认为宇宙是以一种少见的有序方式爆炸的，尽管从大得实际上是必然的概率上看，一次偶然的宇宙创生过程会造成一个全然无序的宇宙。宇宙学的这一基本的悖论引发了好几个不同的反应：

一、理所当然论

很多科学家倾向认为，从一种归纳的基础出发讨论概率、随机性以及可能性是无意义的。假如你在海边随便拣到一块卵石，仔细测量它的尺寸、形状，你就会正确地得到这样的结论：你挑选到具有如此尺寸的卵石的概率极小。但假如进一步说你进行了这样的挑选必定是一个奇迹，或说某种超自然的或神秘的东西引导着你进行了这样的选择，那你可就不对了。因为，你在事后，在拣到这卵石之后再说这样的话是一点也不能令人信服的。当然，假如你所拣到的卵石的尺寸是事先说好的，你是有理由惊奇的。同样我们也可以说，只要宇宙存在，就不必对它特有的结构感到惊奇，因为它就是这个样子。

有一个与此相关的问题是，至少按一种概念看来，概率从定义上讲是与试验的集合相关的。例如，所谓掷骰子掷出“2”来的概率是六分之一，就是说掷过很多很多次后；得“2”的次数差不多是掷的总次数的六分之一。试验的次数越大，得2的次数与掷的总次数的比例就

越接近六分之一这个值。至少，我们就概率所进行的讨论的主题肯定是由一些相似的东西构成的集合的一个成员。例如，骰子的一个面有5个邻面，海边的那块卵石有几百万个面。那么，假如宇宙只有一个，我们来讨论它的可能性又能有什么意义呢？

不过，上面所说的推论不能完全令人信服。假如拣到的那块卵石是完全规则的球形，那么，即使事先没有说好其球形的性质，我们也有理由感到惊奇。因为球形是一种很特殊的形状，它有一个特点，即具有高度的数学规则性。随机地选到一个完全是圆球形的卵石，即使是在事后，也会被认为是罕见的，是应当进行某种解释的。同样，一个适于人类居住的宇宙，对我们这些在绝大多数其他可能的宇宙中不存在的人类来说具有一种特殊的意义：其他可能的宇宙是不能住人的。

对此，持“理所当然”观点的人回答道，假如宇宙当初不是现在这个样子的话，我们也就不会在这里大发惊奇之语了。实际上，任何一个智能生物可在其中提出哲学问题和数学问题的宇宙，不管从演绎的角度看是多么的罕见，也必定是一种我们所观测到的宇宙。换言之，持“理所当然”观点的人认为，我们所感知的高度有序的宇宙并没有什么不同寻常，并不神秘，因为假如它不是现在这个样子的话，我们就不可能（明确地）感知它。

这种推理获得了逻辑实证主义哲学的一些支持。简略地讲，逻辑实证主义认为，谈论我们永远不能观察到的东西是没有意义的。谈论一个其中没有任何有意识的观察者的宇宙有什么意义呢？这样的宇宙永远也不会通过观察被证实或否证，因而，它的存在对有意识的人来说似乎是没有意义的。

与理所当然论相关的一个理论是所谓的强人存原理。天体物理学家布兰东·卡特最先详细地提出了这一理论，近年来物理学家和天文学家对此进行了广泛的讨论。按照这一原理来看，“宇宙必须是这样

的，以便在某一阶段让有意识的生物在其中出现”③（黑点是我标的）这就等于说，宇宙是今天这个样子一点也不奇怪，它没有选择，只能带着适当的秩序出现，以使生命得以产生。

逻辑实证主义和强人存原理这两种理论的成立与否，全系于人类（或天外）智能观察者的至高地位。神学家会说，上帝就是一个观察者，而且上帝的存在不需要特定的物理条件。因此，只要是能被上帝观察到的话，那些永远也不会产生生命的宇宙也是有意义的。

二、多宇宙理论

根据多宇宙理论的观点，有一个由很多宇宙构成的集合，而我们的宇宙只是这个集合中的一个成员。我们所感知的宇宙只是庞大的或许是无限的宇宙集合中的一员。宇宙集合中的每一个宇宙都与集合中其他的宇宙有某种不同。在这集合中会有物质和能量的各种可能的安排。尽管在这集合中的绝大部分宇宙不适于生命存在，而且很接近最大熵的完全混沌的状态（即热力学平衡），然而，在数目极少的宇宙中，偶然地出现了合适的条件，于是生命发展了起来。显然，生物将要感知到的只能是这些偶然的宇宙，而且，这些生物还要写一些书，大谈他们所居住的世界多么不可思议。

上面所提到过的玻尔兹曼的假说，在逻辑上是与多宇宙理论一致的。玻尔兹曼假说中的宇宙是相继发生的，但宇宙获得组织的各个阶段之间有巨大的时间间隔，以致这些阶段在物理上几乎是互不相连的。现代有人修改了玻尔兹曼的宇宙相继发生说，提出了振荡宇宙理论。我们以后就会看到（第十五章），现今的宇宙膨胀可能不会无限地进行下去。假如果真如此，那么，宇宙最终会开始收缩；于是就会出现人们所说的“大崩塌”这样的巨大灾变。有些物理学家推测，宇宙高度收缩之后，并不会缩为看不见的时空奇点，而会在高度质密的状态下“反弹”，从而又开始新一轮的膨胀和收缩。如此看来，宇宙就是这样永无止境地反反复复，一时“大崩塌”，一时再膨胀，进入低密度

状态，颇像是一个不停地充气又泄气的气球。

振荡宇宙是一种年龄无限的宇宙，因而也面临我们在第二章里所讨论的年龄无限长的宇宙所具有的物理难题。然而，围绕着极度塌缩状态这种物理现象的一切不确定的因素拓宽了物理学家们的推测范围。惠勒提出，“大崩塌”具有“重新处理”宇宙的作用。他的意思是，宇宙每一轮新的膨胀和收缩都是一种“新交易”，宇宙所有的物理条件在这交易中被随机地重新凑起来。现在没有谁试图解释这样的事怎么会发生，但假如真有这样的事发生，那么，经过足够多的次数的膨胀收缩之后，宇宙就会经历一遍所有的可能性——当然，所谓足够多的次数必须是一个很大的天文数字。于是，我们再次发现，在一轮又一轮的宇宙膨胀收缩中，只有在偶然搞对了的时候，才会演化出一些宇宙学家来推测宇宙如何创生。

上面所说的一派人的观点是认为在时间中存在着一个由很多宇宙组成的集合，然而，另外有人则猜想宇宙只有一个，这个独一无二的宇宙在空间上是无限的。几乎整个宇宙都接近平衡状态（没有结构也没有组织），但偶然的起伏会造成孤立的有序区域自发地分散出现。当然，这些孤立的有序区域之间的距离远得不可想象，但生命以及有意识的观察者只能在这样的孤立区域中形成，因而，其中的所有观察者就必然会感知秩序。

然而，多宇宙理论的一个或许最广为人知的变体，是由埃弗列特对量子论的解释构成的。在埃弗列特的理论看来，所有可能的量子世界实际上都是实在的，相互平行共存的。因而，一个电子每次面临两个选择时，便会发生两个可能性，于是整个宇宙便一分为二了。一分为二的宇宙当中的每一个都带有所有的居住者（居住者的大脑也一分为二了，而且他们的精神按理说也一分为二了），每一方的居民都自认为电子突然做出了一种选择。这两个宇宙互相分离，一方的居民不能通过普通的空间或时间到另一方去。从某种意义上讲，两个宇宙是“平行”存在的。有多少量子选择，就有多少宇宙，因而，在无穷的

平行世界当中，一切可能的物质和能量的排列都会发生。

观察者从数目庞大的多种选择中选出一个高度非典型的宇宙，这种理论被称作弱人存原理。有人根据若干哲学和物理学的理由对这一理论进行了驳难。首先，从某种意义上讲，这个理论过于成功。这种理论认为大自然能够使一切可能实在化，这样，一切就可以都得到“解释”了。实际上这样一来，我们就可以不要科学了。只要说明某某事物对人的存在是必需的，于是，它一下子就算是得到解释了。

人择原理所受到的另一个驳难是，它似乎是与奥科姆剃刀原理正相反。因为根据奥氏原理，在一套可能的解释中，最有道理的是那个包含的原理最简单而且假设又最少的解释。求助于无限多的宇宙来解释一个宇宙显然是太过分了，过分到宇宙规模上了，且不说那无限多的其他宇宙除了极少数以外都从未被观察到（或许只被上帝观察到了）。持人存原理观点的人反驳道：“根本没这回事。埃弗列特的量子论解释或许牵涉的宇宙是多一些，但在认识论上是极其简洁的。想一想吧，对量子测量问题的其他解释是多么牵强，多么没有道理。而在多宇宙理论中，解释只是来自形式主义，不需要另外的形而上的假说。”

然而，持多宇宙理论的人承认，他们的理论中的“其他的世界”甚至在原则上也永远不可能被观察。在分岔的量子世界之间往来是不可能的。而且，无限多的或振荡的典型宇宙中有序的区域彼此间隔着如此之大的空间或时间，以至没有哪个观察者能从经验上证实或否证多宇宙的存在。人们难以明白，这样一个纯粹的理论框架怎么能在科学的意义上被用作一个自然的特征的解释。当然，人们或许会觉得，比起相信一个无限的神明来，相信宇宙的数目无限要容易一些，但这样的信念只能以信仰而不是以观察为基础。

弱人存原理和强人存原理二者的科学基础也受到了质疑。整个的人存原理的基础是概率的概念，然而有人也利用概率的概念来反驳

它。这里的问题涉及小起伏对大起伏的相对可能性。我们可以再想一想那个乱弹琴的黑猩猩。那黑猩猩乱弹很长时间后，我们有理由期望听到一个熟悉的曲调的一串3个或4个音符。假如要想听到黑猩猩弹出有6个音符的乐句，等的时间就要长得多。随着有序程度的提高，黑猩猩弹出正确的音符的可能性便陡然下降。再举一个例子。4个人摸一副洗过的扑克牌，很可能每人都摸到1个A。但是，每个人都摸到1个A和2、3张同花的顺牌的可能性就小了。而每一个人都各摸全一套同花牌的可能性则极小极小。这是因为，小的巧合相对而言要比大的巧合可能性大得多。

从宇宙学上看，一个随机事件造成一颗恒星的几率比起造成整个一个星系的几率要大得多。而随机事件造成几十亿星系的可能性比起造成一个星系的可能性来，就该是无穷小了。但是，据有人推理，确实只有一个星系——很可能只有一个恒星适于生命形成并出现观察者。这推理对吗？那么，为什么我们观察到整个宇宙充满了结构？多宇宙理论认为，大多数宇宙都只有一个星系，每一个具有两个星系的宇宙，便有无数个单星系的宇宙与之对应。假如有更多的星系，比例的差异就迅速地增长。假如所有的宇宙当中都有观察者，那么，他们当中的绝大多数是居住在单星系的宇宙之中，而不是居住在多星系的宇宙之中。那么，我们又怎样解释在我们的宇宙中存在着如此众多的星系呢？

对此，人们所能想出的唯一回答是，由于某种尚未明了的原因，一个星系的形成不知如何与宇宙的大尺度结构联系了起来。很可能，只有在某种特定的整体条件具备的时候，星系才能够形成；而一旦出现了这种特定的条件，星系就在各处得以形成。换言之，宇宙要么到处有星系，要么就到处都没有星系。星系的形成与宇宙大尺度结构这种总体联系在物理学上现已明了，但星系形成的机制现仍很不清楚，不能据以对星系形成的可能性作实际性评价。

三、有序来自混沌论

对宇宙秩序的起源之谜，第三个反应是有人试图证明，宇宙的秩序是通过自然的物理过程从起始的混沌状态中产生出来的（这里所说的“自然的物理过程”不仅是罕见得不可想象的起伏）。（在第四章里，我们详细地讨论了这种理论，所以在这里只作一个简短的总结。）乍看之下，这种理论似乎注定要失败。热力学第二定律岂不是说了吗，秩序能走向混沌，而混沌则不能走向秩序（这里暂且不谈偏离平衡态的起伏）。

事实的确如此，但我们还得把热力学第二定律搞得细一点。严格地说，热力学第二定律只是为应用于完全封闭的系统而提出的。显然，任何一部分宇宙，不管有多大，也不是封闭的，因为它与周围的那些部分的宇宙有接触。更为重要的是，整个宇宙都在膨胀，这是大家都知道的。而膨胀这种外在的扰动会造成完全出人意料的情况。

这里可以举一个很好的类比。设有一个普通的汽油发动机上的活塞和汽缸，一种气体被活塞堵在汽缸中。假如活塞处于静止状态，该气体便处于平衡状态，在汽缸中各处的温度和压力都是一致的。这是一种熵值最大的状态。在这种情况下，我们不能期望汽缸中的气体会进一步发生变化：该气体没有任何有序的结构或有组织的行为。假设现在活塞被猛然提起，使气体膨胀，那么，这种气体便立刻不再到处都有同样的温度和压力了。靠近退去的活塞处的气体密度变小了，因为它的空间大了。当气体流向这一空间时，便发生了湍流。假如活塞又被推回来，推到起始的位置上，气体就又会静下来，进入新的热平衡状态，但由于这种扰动，熵会增大。当活塞运动时，气体会暂时生出一种结构和组织。

在热力学第二定律中，我们有没有发现漏洞呢？没有。气体的熵在活塞运动一周之后增大了（气体更热了）。起始的平衡状态曾是最大熵的状态，这状态与对这活塞气缸系统的外在拘束是一致的。然而，当活塞运动时，这些外在的拘束发生了变化，使气体得以寻求更高熵的状态。简言之，起初的平衡状态只是一种相对的最大状态，而

不是绝对的最大熵状态。

在宇宙学的场合中，宇宙的膨胀所扮演的角色与上面所说的活塞类似。宇宙膨胀和活塞活动都使外在的拘束发生了变化。宇宙学家们指出，原初宇宙根本就不是处于一种有序的状态之中，而是接近于热平衡。我们现在所观察到的各种常见的结构——星系、恒星、原子——在大爆炸时都不存在。实际上，在宇宙创生之初大约一分钟之内，温度是如此之高，以至连原子核也不能存在。但不知为什么，从原初的混沌中产生出宇宙现在的有序结构。到底是如何产生的呢？

我们所熟悉的地球上的大多数复杂组织，如生态系统和天气模式，都是由太阳光产生的。太阳光是最重要的负熵来源，地球上一切复杂的组织都赖其为生。太阳的负熵储藏是其核燃料（主要是氢）。核物质自身最高熵形态是由质量不大也不小的元素构成的，如铁元素。阳光的产生就代表太阳产生的熵。太阳试图通过一系列核反应将氢变成铁，于是就产生了阳光，产生了熵。太阳有序（负熵）的秘密，以及大多数其他恒星保有负熵的秘密，应当在其氢含量中寻找。宇宙中的物质大约有四分之三是氢，剩下的四分之一几乎全部是质量之小仅次于氢的物质——氦。为什么这四分之一不全是铁呢？

关于这个问题，第四章已给出过答案。原初的宇宙当时太热，容不得铁存在，随后，宇宙冷却得太快，容不得足够数量的核反应发生。于是，原初物质就这么落在低熵的氢的形态之中，不能够达到变成高熵的铁的目标。只是后来出现了恒星，这种情况才发生了变化。

按着这一思路解释，显然就不必假想宇宙在创生之初就处于显著的有序状态之中。原初的物质实际上是处于一种完全无序的状态（最大熵）。这样的状态可以通过很多很多的方式来达到，带着别针的造物主只要随便将别针往“采购单”上一别就行。宇宙有序的起源之谜这就算解开了。

难道真是解开了吗？

的确，宇宙物质的核子状态在产生我们所观察到的结构和组织的过程中是一个至关重要的因素，但它还不能将宇宙的一切都解释出来。那些较大的结构——恒星和星系——是由引力造成的。而且，至关重要的宇宙膨胀也是受控于引力的。关于宇宙的引力组织，我们又知道些什么呢？从引力的观点来看，我们生活于其中的宇宙是高度有序的，还是无序的？下一章的主题，就是这些问题。

# 第十三章　黑洞与宇宙混沌

混沌是无处不在的。

约翰·巴罗

我们的宇宙在创生之初就处于一种非常特殊的状态吗？它是经过精心塑造，以便在一定的时间之后先让生命、最后再让精神繁盛起来惊叹宇宙的奇妙吗？若不是这样的话，那么，我们是生活在荒唐至极毫无意义的偶然事物之中吗？我们的宇宙真是从虚无中爆发出来的吗？宇宙的创生真是纯粹的偶然事件吗？的确，现今的宇宙学家所要对付的最为迫切的中心问题，莫过于这个有关存在的问题了。

在上一章里所讲的诸种论点表明，尽管热力学第二定律具有不可抗拒的规定，但宇宙秩序在很大程度上能够自然地毫不奇怪地从完全是混沌的原初宇宙中产生出来，而原初宇宙完全是偶然的，代表着物质世界的偶然起源。然而，假如把引力考虑进来的话，宇宙秩序的起源的图景就大大不同了。

引力是大自然的各种力当中最弱的，但其力量具有累加性，所以在大尺度上引力占了优势。要解释星团、星系的结构以及膨胀的宇宙的总体运动，我们就得仰仗引力。尽管爱因斯坦的广义相对论中的空间弯曲和时间弯曲使人们相当理解引力的性质了，但说到引力秩序，物理学就乱了套。关于受引力作用的系统的热动力学，目前尚没有一致的看法或理解，诸如引力场的熵之类的概念现在仍是模模糊糊，没有清晰的表达。

在第四章里说过，引力熵的一个具有悖论性质的方面是，那种在我们看来更有结构的状态，实际上其熵比不那么有结构的状态的熵要高。例如，起初均匀分布的恒星会松弛为一种更为复杂的组织，运行速度快的恒星密集在引力的中心附近，而运行速度较慢的恒星则散布在外围（见图7）。受引力作用的系统会自发地生出结构这种倾向，就是自组织的一个好例子。应该把这个例子与在引力可忽略不计的情况下的气体行为作一番对比。在引力极微，可以忽略不计的情况下，气体便倾向进入一种均匀状态，各处的温度和密度相似。然而，受引力作用的系统则会集结，变得不相似。

在其他的各种力不在的时候，所有的受引力作用的系统都会完全崩溃。例如地球之所以没有被其重量压垮，完全是倚仗构成地球的物质的硬度（这硬度的根源是电磁力）。同样，太阳之所以没有坍缩，也只是因为太阳核中的核反应产生了巨大的外向压力。假如除去地球和太阳内部的这些力，地球和太阳便会在短短的几分钟之内收缩，越缩越快。随着收缩的进行，它们的引力就会上升，收缩的速度也就加快。很快地，它们会被吞没在逐步上升的时间弯曲里，变成黑洞。从外面看上去，时间就好像是停止了，不会看到进一步的变化了。黑洞所代表的，就是受引力作用的系统的最终平衡状态，这状态相当于最大的熵。尽管一般的受引力作用的系统的熵尚不明了，然而，雅各布·贝肯斯坦和史蒂芬·霍金利用量子论对黑洞进行的研究工作却得出了有关这些系统的熵的一个公式。正如所预料的那样，黑洞的熵要比同质量的一个恒星的熵大得多。假定熵与概率的关系可用于受引力作用的系统，便可以用一种很有意思的方式来表达根据该公式得出的结果。设有些随机分布的（受引力作用的）物质，可能性极大的情况是，这些物质将会形成一个黑洞，而不会形成一个恒星或一块由离散的气体构成的云。因而，考虑到这种情况，我们对宇宙是创生于有序还是无序状态这一问题就会有一种观点。假如宇宙的初始状态是随机选择的，那么，大爆炸生出黑洞而不是生出离散气体的可能性就极大。现在的物质和能量的排列是，物质以恒星和气云的形态相对稀薄地分

布，这显然只能是由对宇宙初始状态的非常特殊的选择造成的。罗杰·彭罗斯曾计算过我们所观察到的宇宙出现于偶然的可能性，得出的结论是，从演绎的角度来看，偶然事件造成一个黑洞宇宙的可能性比造成现有的宇宙的可能性要大得多。他估计的数字是$10^{1030}$：1①。

我们所观察到的宇宙不是黑洞（至少不是以黑洞为主），这暗示着宇宙的初始不是偶发事件造成的。不仅如此，宇宙大尺度的结构和运动也同样不同寻常。宇宙的累积引力使宇宙的膨胀受到抑制，使膨胀随着时间减速。在初始阶段，宇宙膨胀的速度要比今天大得多。大爆炸造成了向外的冲力，而引力又使劲要把爆炸的碎片向后再拉到一起去，于是，这两种力量的较量就造成了宇宙。近年来，天文物理学家们才意识到这较量是多么微妙地保持了平衡。假如当初大爆炸的力量不那么强，宇宙便会很快在一场大崩塌中缩回去。然而，假如当初大爆炸的力量更强一些的话，宇宙物质便会如此迅速地离散开去，星系就不会形成了。无论怎样说，宇宙现今的结构似乎非常微妙，全系于大爆炸造成的外冲力和向内的引力的精确平衡。

这平衡到底有多么微妙已由计算揭示了出来。在所谓的普朗克时间（即$10^{-43}$秒，这是时空概念具有意义的最初的时刻），这平衡精确到了$\frac{1}{10^{60}}$。这就是说，假如初始时的爆炸力有$\frac{1}{10^{60}}$的差异，我们现在所观察到的宇宙便不会存在。为了玩味一下这个数字的意义，你可以设想你要从可见宇宙的另一端，也就是从200亿光年以外的地方，射中一个1英寸的靶。你若想真的射中，瞄准的精度就得要精确到$1/\frac{1}{10^{60}}$。

除了这总体的精确性之外，还有一个未解之谜，就是为什么现今的宇宙无论在物质分布上还是在膨胀速度上这么均匀。大部分爆炸都是没有秩序的，人们理应预期大爆炸的冲击力会处处不同。但事实并非如此。我们这边宇宙的膨胀速度与另一头的速度没有分别。

当我们考虑到所谓光视界时，整个宇宙行为的这种一致性似乎就

更加不同寻常了。当光在宇宙中散布开去时，它必须追赶因宇宙膨胀而向后退去的星系。一个星系的退行速度要视该星系与观察者的距离多大而定。遥远的星系退行得要快一些。现在，我们设想在宇宙创生的那个时刻，一束光从某处射了出来。到现在为止，这束光行程大约已200亿光年了。200亿光年以外的那些宇宙区域现在尚未接收到这束光。在那些区域的观察者将看不到该光的光源。反过来说，靠近光源的观察者也看不见这些区域。于是，宇宙之中没有哪个观察者现在能够看到200亿光年以外去。空间之中有一种视界，它把在它之外的一切都掩藏了起来。因为没有哪种信号或影响传播得比光还快，于是，宇宙间各处于彼此的视界之外的那些区域之间就不可能存在任何的物理联系。

当我们把望远镜指向可见宇宙的外围区域时，我们所探测的显然就是彼此间从未有过因果联系的宇宙区域。这是因为，那些从地球上看去是天各一方的区域相距是如此遥远，以至彼此都在对方的视界之外了。这种情况很类似于日常生活中的地平线的情况。海上的一只船上的瞭望员可能正好刚能够看到一前一后的两只船在靠近地平线的地方，但这两只船却彼此看不见，因为它们相距过远。同样，天各一方相距遥远的星系也这样彼此处于对方的视界之外。因为一切物理影响或通讯都是受光速限制的，所以，这些星系是不可能协调它们的行为的。

我们刚才说到的未解之谜是，为什么宇宙这些没有因果联系的区域在结构和行为上这么相似？为什么它们的星系平均大小以及形状都相同？而且为什么它们以同样的速度彼此相离而去？这种协调行为在星系最初形成的时候就有了，当我们意识到这一点时，上述的未解之谜就更意味深长了。但是在过去，自宇宙创生以后，光的行程不是那么远，因而各视界也不那么大。宇宙创生100万年以后，各视界就是100万光年，宇宙创生100年以后，各视界就是100光年，依此类推。假如我们再回到普朗克时间，各视界的大小就只有$10^{-43}$厘米了。即使

把宇宙的膨胀考虑进去，按照标准理论来看，这些如此之小的区域到现在也膨胀不到可以看到的尺度。现在看来，当时整个可观察到的宇宙似乎是分成至少$10^{80}$个彼此没有因果联系的区域。怎样才能解释这种没有联系的协调呢？

与此相关的另一个难题是，宇宙具有极高的各向同性的性质，即方向上的均匀性。从地球上向外看，宇宙在大尺度上是一样的，不管我们朝什么方向看都是一样。仔细测量残留的宇宙背景热辐射之后，人们发现从各个方向来的辐射通量具有精确的均衡，相差仅千分之一多一点。假如大爆炸是一个随机事件，这种异乎寻常的均匀性几乎是绝对不可能的。

考虑过这些情况之后，人们得出的结论是，宇宙的引力排列之规则、均匀是令人迷惑不解的。似乎不存在任何明显的理由能够说明，为什么宇宙没有乱了套，没有以无秩序无协调的方式膨胀开去产生众多的黑洞。把大爆炸的巨大威力导入这样一种有规则有组织的运动模式，这似乎是个奇迹。难道真是奇迹吗？我们现在来仔细探讨一下对这未解之谜的各种反应。

一、隐藏的原理

当某个量被发现具有一个非常接近零的值时，物理学家们就倾向于猜想，因为某种深刻的原因，其值就是零。他们就会去寻找某种基本原理，好据以确认该量肯定恰好为零。例如，不同的电子所携带的电荷没有可辨别得出来的差别。于是，物理学家们就得出结论说，那些电荷是恰好相同的，即它们之间的差别为零。这就是由电子的不可分辨性的基本原理导出的推论。另一个例子是，一切一同下落的物体都一同触地（在没有阻力的情况下）。它们到达地面的时间差别被看作正好为零，这就是所谓相等原理的推论。这个引力的基本原理规定，一个物体对引力做出的反应与该物体自身的性质无关。

我们可以设想有一原理（或一套原理）规定，大爆炸的冲击力正好处处与其引力相称，于是退行的各个星系正好摆脱了它们自己的引力。这就意味着，宇宙的膨胀方式恰好处于那个分界线上，一方面是宇宙物质的彻底离散，另一方面是宇宙停止膨胀，接着崩溃。这样的一个原理也可以保证，宇宙从大爆炸中降生时就具有分布均匀的物质，而不是爆炸出些黑洞。同样，这样的一个原理能够保证，宇宙的膨胀在所有的方向上都是恰好均匀的。尽管我们一点也不知道这些原理会是什么样子，然而，当我们知道宇宙在其各个不同区域及方向上的膨胀速度的差异非常接近于零时，我们就难免这样认为：有一个大自然的原理使得这些差异必定恰好为零。

不幸的是，事情不可能这么简单。假如宇宙果真是完全均匀的，那也就不会形成任何星系了。照人们现在的理解来看，星系从原初的星云中形成似乎只能形成在自宇宙创生到现在的这段时间里，假如星系的雏形一开始就有的话。星系从其周围的宇宙环境中积累物质积累得很慢，而宇宙膨胀的速度却相对很快。星系只有早一点起始，才能克服宇宙膨胀带来的离散倾向。假如真有一个基本原理，那么，似乎该原理必定允许恰好充足的偏离均匀状态的情况发生，使星系得以生长起来，同时却不产生出黑洞。这确实是个微妙而复杂的平衡行为！

二、耗散

宇宙的均匀膨胀也可能这样解释：宇宙起始时具有高度非均匀的运动，但不知何故，宇宙把这种非均匀的运动耗散掉了。理论研究确实表明，假如一个宇宙在一个方向上的膨胀比在其他方向上快得多，那么，它就会因多种机制而被刹车。例如，从膨胀的能量中创造物质（见第三章）就会耗尽膨胀速度快的那一方向上的运动力量，使之逐渐与其他方向一致起来。人们还知道有其他的刹车过程。

对这一观点，有人提出了两个反驳。第一个是，不管对原初非均匀运动的耗散是多么有效，总是可以发现宇宙的起始状态的，因为其

起始状态是那么不均匀，即便是有衰减，也还会留下残迹的。人们充其量只能证明，宇宙肯定起源于一群不同寻常的起始状态。

第二个反驳是，一切消耗都产生熵。宇宙原初剧烈的非均匀运动会转化为极大量的热，大大多于我们所观察到的原初热辐射的量。然而，这一反驳有一个漏洞。因为它所说的宇宙中的热量是一个无意义的概念。要想使之有意义，就必须用某种尺度或标准来测量它。而唯一能够找得到的比较标准是物质。于是，宇宙学家们就采用了每原子热量，或更精确地说是每质子热量这一概念。这就是说，他们计算出在某块大体积的空间内的热量，然后，再估计在这块空间体积之中的物质质量，计算出相应的质子数，从而得出单位质子的热量是多少。结果发现，单位质子的热量很小。要想使之与平衡的热量输出相符，就得将它扩大几乎$10^{15}$倍。我们在这里所讲的对耗散说的第二个反驳认为，单位质子的热量之所以小了这么多，是因原初宇宙的沉寂性质所致。假如原初宇宙当初确实骚动不安，具有高度的非均匀运动，那么，现在的空间就会充满了把人烤出油来的热辐射。但是，漏洞就出在用质子来测量热值上。质子是可以破坏的，这样的粒子不能够充当固定的比较标准。根据所谓的基本力的大统一理论，质子是可以衰变的。而且，质子也可以（通过与衰变逆向的过程）被创造出来。我们在第三章里说过，质子如何被从原初能量中创造出来，也说过大统一理论用其参数预测（实际上这预测是正确的）出单位质子的热量。因为大统一理论会自动地调节质子丰度，使之与实存的热量相合，所以，不管宇宙原初的非均匀运动消耗了多少原初热量，大统一理论最后得出的单位质子的热量也是相同的。因而，宇宙究竟起始于沉寂的、高度均匀的状态，还是起始于极端非规则、非均匀的状态，这一问题的答案要由将来是否能够证实大统一理论（可能是通过证实质子衰变的方式）来定。

三、人存原理

假如一个宇宙充斥着黑洞，或充斥着大尺度的非均匀运动，那

么，这样的宇宙是不可能有助于生命形成的。于是，现今的宇宙的均匀性就显然给人存原理留出了地盘。假如用弱人存原理解释宇宙的均匀性，我们就可以设想有一个由无数个宇宙构成的集合，包含了宇宙起始膨胀运动以及物质分布的所有的可能选择。生命和观察者只能形成于极少数的宇宙中，这些宇宙的排列接近于我们现在所观察的宇宙。各向异性的或高度非均匀的宇宙是不可认识的。

要使这一解释成为一种成功的解释，就得证明宇宙的非规则性增加哪怕一丁点也会对生命有害。很可能，原初状态发生微小的变化，都会使宇宙现在的物理环境面目全非。例如，假如质子不会衰变，那么，原初的宇宙一点点各向异性也能产生大量的热，使生命不可能存在。宇宙的背景温度仅增加100倍，也会造成我们所知的生命的毁灭。然而，现在尚未有人对此进行详细的计算，因而人存说就容易受到我们在前面一章里所说过的批评。

四、暴涨

有人在最近对宇宙的均匀性提出了一种全新的解释。这种解释源自大统一理论，然而，其至关重要的立论基础却是若干有待商榷的关于超高能物质的假设。这些假设在任何情况下都是难以证实的。不过，这一解释生动地说明，基本物理学的进展可以改变我们对宇宙秩序起源的整个看法。

我们还记得，当宇宙从大爆炸冷却下来的时候，大自然的三种基本力——电磁力、弱核力和强核力——由当初没有分别的状态被“冻结”成现在的不同形式。这个从一种状态变到另一种状态的转换近似于水蒸气变成水或水变成冰。两种状态不仅有力的性质的差别，而且它们的引力作用也有所不同。把那大统一力分裂为不同的电磁力和核力的机制，也产生了巨大的推斥性引力。

某种宇宙斥力存在的可能性，实际上是爱因斯坦在1917年提出

的，尽管爱因斯坦本人实际上一直不喜欢这一概念，而且现在在天文学上也没有任何证据显示有这样一种力存在。不过，大统一理论认为，在灼热的宇宙原初阶段，宇宙斥力的存在必然是不可避免的。那是约在宇宙年龄为$10^{-35}$秒之前，当时，宇宙温度高得不可想象，为$10^{28}$K。麻省理工学院的艾伦·古斯指出，宇宙斥力的存在会对原初宇宙的结构产生巨大而深远的影响。

当宇宙膨胀并冷却下来的时候，斥力似乎很可能压过通常的具有吸引性的引力，致使宇宙开始进入无法控制的剧烈暴涨状态。在极短极短的时间里，一块极小极小的空间区域会以指数的方式膨胀至宇宙大小，大约每$10^{-35}$秒体积就增加1倍。这种势不可挡的膨胀会一直持续到某一时刻，那时，宇宙就会突然进入另一个“冻结”状态，在这种状态中，大统一力分成不同的力，斥力不复存在。在没有斥力这一股强大的力存在的情况下，空间的指数增长就会在剧烈的放热中停下来，宇宙就会进入更为正常的缓慢减速膨胀，而今天，这种减速膨胀活动的残迹仍然存在。

暴涨宇宙理论一举解决了好几个主要的宇宙学难题。例如，它说明了为什么宇宙如此均匀。任何起始的不规则都会因这种巨大的暴涨而被大大地“冲淡”。一块体积不过质子大小的空间可能会暴涨至现在可以观察到的宇宙体积的很多倍。因而，宇宙中的质子尺度或质子尺度以上的不规则性就会在我们可以观察到的宇宙中被拉长至微不足道的程度。

暴涨理论也能解释大爆炸的外冲力与宇宙物质的引力之间奇迹般的平衡。古斯认为，来了指数暴涨时，宇宙膨胀速度的过量或不足就被消灭了，这同时阻止了可怕的黑洞在宇宙原初阶段形成。到了宇宙从那指数暴涨里退出的时候，宇宙膨胀速度的过量或不足就会被降至非常接近于零。（尽管显然不恰好为零，因为星系现在仍能形成。）

最后，暴涨也解决了视界问题。天各一方的宇宙区域，通常被看

作是没有因果联系的，然而，它们实际上在暴涨阶段之前有过短暂的因果联系。我们所观察的一切（还有很多我们观察不到的东西）在暴涨开始时，都挤在狭小的空间区域里。视界并不存在（至少不存在于我们以前所想的地方）。预言视界的存在，是基于假设宇宙创生以来的膨胀一直是平稳地减速，而没有考虑到指数增长的时期。

暴涨说干净利落地说明了宇宙学的几个由来已久的问题，然而，这一学说也不是没有滞碍难通之处。其主要滞碍是所谓的“优雅退场”问题。暴涨若要现出其魔术，那么，指数增长的时间就必须足够长，以使宇宙膨胀10的很多次幂。宇宙突然大规模地膨胀，使温度几乎立刻降至很接近于绝对零度。在我们看来，似乎没有什么能来阻止“冻结”即时发生，这样一来，暴涨还没有正式开始就被阻挡住了。

古斯在早期的暴涨理论中提出，宇宙很可能经历过一个所谓过冷的时期。物理学家们知道在日常生活中有这种过冷现象。例如，假如水是纯净的，便可以仔细小心地把它冷却到冰点以下而不结冰。然而，假如稍有扰动，过冷的水就会突然变成冰。而宇宙的过冷能够使宇宙在高温（统一力）阶段停留足够长的时间，使暴涨得以进行。但冻结发生时便来了问题。新的（“冻结的”）状态的“小泡泡”似乎可能会随机地出现，而且会开始以光速生长。在这些小泡泡里面没有暴涨，因为暴涨式生长的能量被转移到泡壁上了。最后，这些小泡泡会大到相互交叉的程度。具有高能的泡壁相互碰撞便会引发大规模的非均匀和非规则的东西，而这些非均匀、非规则的东西恰恰是暴涨说本来要解释的东西。

物理学家们仍在继续进行研究，以避开这乱麻一样的难题，使之不再困扰暴涨理论。有人提出一个见解，认为冻结状态的小泡泡会在暴涨中变得如此之大，足以包容整个的宇宙和很多其他的东西，因而，尽管在很大的尺度上，宇宙是不规则的、骚动不安的，但我们所观察到的宇宙则是其中的一块相对均匀、静止的区域。还有人提出，过冷阶段过去之后不是随即出现冻结状态的小泡泡，冻结过程很可能

是慢吞吞的，这样，就能赶在宇宙从过冷状态转入冻结状态之前，留出相对长的时间使暴涨得以发生。这一理论的很多细节是高度依赖模型的。现在尚不能说，优雅退场的问题是否会得到满意的解答。

尽管有些技术性的滞碍难通之处，暴涨说所取得的广泛成功赢得了很多物理学家和宇宙学家的喜爱。假如这一学说是正确的，那就是说宇宙不必以一种非常特殊而有序的状态被创造出来。起始的引力不规则被暴涨消灭了，而随后而来的宇宙膨胀又使起始时无结构的宇宙物质演化出复杂的组织和结构。因而，复杂的宇宙秩序的起源完全可以被解释作纯自然过程的结果。

五、上帝

假如大统一理论不能说明问题，假如人存原理不被人接受，那么，宇宙大尺度的高度均匀的性质可被提出来作为一个具有创造力的设计者存在的证明。然而，这证明只能是消极的。谁也不敢保证，随着我们对早期宇宙的物理规律的认识进步，我们不会发现一种完满的解释来说明宇宙的有序。早先，太阳系复杂而有序的结构曾被认为是神造的，但后来被归入标准的天体物理学的领域，同样，大尺度宇宙秩序的那些未解之谜也可能有一天会得到自然的而非超自然的解释。

我们的结论必定是，现在没有肯定性的科学证据，证明有一个宇宙秩序（即负熵）的设计者和创造者存在。实际上，人们有把握地预期，现有的物理学理论将会为那些未解之谜提供完全令人满意的解释。

然而，大自然除了数学规律和复杂的秩序二者之外，还有更多的东西。大自然的第三个成分也需要解释。这第三个成分就是所谓的大自然的“基本常量”。我们正是在基本常量这一领域里，发现了一个宏伟设计的最为令人震惊的证明。

物理学家们所说的基本常量，是指一些在物理规律中扮演着基本角色的量，这些量在宇宙中所有的地方，在时间的所有的时刻里，都有相同的数值。用几个例子就足以说明这一点。在一遥远的恒星上的一个氢原子与地球上的一个氢原子是完全一样的。那里和这里的氢原子，其大小、质量以及内部的电荷都是相同的。但是，这些量的值在我们看来完全是一个谜。为什么氢原子里的质子重量是电子的1836倍？为什么会有这个数字？为什么它们的电荷是现在这个值，而不是某个其他的值？

大自然的各种基本力都包含着这种决定它们的强度和范围的数值。或许在将来的某一天我们将会有一个理论，可以以一种更基本的概念来解释这一切数字。不管怎样，反正人们发现这些常量的值对物质世界的结构具有关键性的意义。

我们且来看一看弗里曼·戴森所提出的一个简单的例子吧。强核力将原子核胶结起来不致散开，而强核力来自我们在第十一章里所讲的夸克和胶子。假如强核力不像现在这么强，原子核就会变得不稳定，就会蜕变。最简单的复合原子核是氘（重氢），是由一个质子附在一个中子上构成的。这中子和质子被强核力胶粘起来，但粘的并不牢固。假如核力弱百分之几，质子和中子的结合就会被量子分裂所破坏，结果就会是戏剧性的。太阳以及大部分恒星都把氘用作连续核反应的一个环节，以便持续不断地放光。假如除去氘，所有的恒星则要么熄灭，要么就再找一条进行核反应的途径来继续产生热。熄灭也好，再来一条新途径也好，都会大大地改变恒星的结构。

假如核力比现在稍强一点点，也会发生同样糟糕的后果。核力若是再稍微强一点，两个质子便有可能克服它们共有的电斥力而粘在一起。在大爆炸期间，质子要比中子多得多。当原初物质冷却下来时，中子就寻找质子以与之粘合起来。质子与中子的结合构成了氘，氘很快又经过进一步的结合，形成了氦元素。但是，剩余的质子完好如初，构成了制造恒星的原料。假如这些质子可以成对地粘合在一起，

那么，每一质子对当中的一个质子就会衰变为中子，使质子对变为氘，然后再变为氦。这样，在一个核子力强出百分之几的世界里，几乎就不会有氢从大爆炸时留下来。类似太阳的稳定的恒星就不会存在，液态的水也不会存在。尽管我们不知道为什么核力具有现在的强度，然而，我们却知道，假如核力不是现在这样的，宇宙就会面目全非。生命是否能够存在也就成了问题。

使很多科学家觉得非同寻常的，倒不是基本常量值的改变会改变物质世界的结构，而是人类所观察到的物质世界的结构对常量的改变很敏感。基本力的强度稍微有一点差异，就会造成物质世界结构的剧烈变化。

还可以再考虑另外一个例子，是有关物质的电磁力和引力的相对强度的。电磁力和引力在形成恒星的结构的过程中，扮演的角色十分重要。恒星是由引力粘结起来的，引力的强度辅助决定了诸如恒星内部的压强之类的事情。另一方面，能量又以电磁辐射的形式从恒星中逸出。引力和电磁力这两种力的相互作用是复杂的，但个中的道理还是能被人很好地理解。一般说，重的恒星较亮，较热，可以毫无阻碍地将星核发生的能量以光和热辐射的形式输送到星的表面。轻的恒星则较凉，其内部仅以辐射的方式不能够足够快地放出其能量，因而必须辅之以对流，这就使它们的表层发生沸腾。

这两类星——热的、辐射的和凉的、对流的——分别被称作蓝巨星和红矮星。这两类星为恒星的质量划定了一个很窄的范围。而实际情况是，恒星内部电磁力和引力的平衡竟然确实使几乎所有的恒星都处于蓝巨星和红矮星之间的狭窄范围之内。然而，正如布兰东·卡特所指出的，②之所以会有这样幸运的事，完全是因为大自然基本常量之间具有奇妙的数值巧合。比如说，假如引力的强度变化哪怕是$\frac{1}{10^{40}}$，也足以破坏掉这种数值巧合，于是，所有的恒星便只能不是蓝巨星就是红矮星。像太阳这样的恒星就不会存在，而且我们还可以说，依存

于太阳一类的恒星的生命也不会存在。

我们所观察到的世界的结构之所以存在，完全是因为有这样一些数值巧合，但这些数值巧合的数目太多，在这里不能一一评论。（读者若想了解有关的完整讨论，请看我的一本书《偶然的宇宙》。）对于这些数值巧合，物理学家们见解不一。宇宙的初始条件好像是经过设计，于是，为了解释这一点，有人就可以求助于人存理论，假设存在着很多宇宙，在这些宇宙中，基本常量因某种原因具有不同的值。只有在那些基本常量恰好合适的宇宙中，才会形成生命和观察者。

基本常量的数值巧合也可被看作是设计的证据。常量的数值十分微妙，没有它们，物理学的各个不同的分支就不会接合得如此巧妙。这可被看作是上帝的所为。宇宙现今的结构显然对基本常量数值的微小变更十分敏感，因而，人们很难不这样想，宇宙的结构是经过精心设计的。当然，这样的一个结论只能是主观的，最终要归结为一个信仰问题。相信有一个宇宙的设计者来得容易一些呢，还是相信存在着弱人存原理所要求的多宇宙容易？很难设想，这两种假说能得到严格的科学验证。正如前一章所指出的，假如我们不能探访其他的宇宙或不能直接地感受它们，那么，相信它们存在就同相信上帝存在一样，必定仍是一个信仰问题。或许，将来的科学发展会使人们找到其他宇宙存在的更直接的证据，但是，在那天到来之前，大自然赋予基本常量的奇迹般的数值巧合就必定仍然是宇宙经过设计的最不可抗拒的证明。

# 第十四章　奇迹

上帝从未制造奇迹来说服无神论，上帝平常的工作就足以说服无神论了。

弗朗西斯·培根

任何奇迹，在历史上都找不到足够多的具备毫不可疑的判断力、教育和学识以使我们相信其不受迷惑的人来加以证实。

大卫·休谟

某些关于上帝存在的论证，其根基是宇宙学和大自然中存在着设计的想法。不管这样一些论证多么具有说服力，它们充其量也是间接的。然而，有些人声称，可以通过奇迹直接在物质世界中看到上帝的作为。世界所有的大宗教都有关于奇迹的民间传说。圣经里就有很多关于奇迹的叙述，甚至时至今日，仍常有奇迹的报道。

假如我们想评价奇迹的证据，碰上的第一个问题就是确切地断定“奇迹”是什么意思。然而，奇迹的意义是什么，根本就没有一致的意见。“现代科学的奇迹”，意思是某种非同寻常的、壮观的东西，但在说这话的时候，没有人认为其中的“奇迹”一词用的是其本意。阿奎那把奇迹定义为某种“由神力脱离事物的常规做成的”东西。用现代语来解释，就是上帝违反大自然的规律的作为。换言之就是，上帝直接插手于世界的运行过程，用“破坏大自然的一切规则”的方式改变了某种东西。假如这类的事件能够得到确证，那么，上帝的存在以及上帝对这个世界的关注也就确实得到了有力的证明。

不过，有时奇迹一词的意思也不是这么重的。一个幸运的人在“奇迹”中多次大难不死，于是便坚信了上帝的仁慈。飞机失事之后的那位唯一的幸存者，可能会把自己的幸免于难看成是一个奇迹，尽管同是一个奇迹，却使所有同机的人无缘无故地死去了。

大难不死必有天助意义上的奇迹，与明确违反大自然规律意义上的奇迹分属完全不同的范畴。没有谁认为，在空难中不死肯定是因为物理定律暂时失效。大难不死之类的事件不过是些令人惊讶的巧合而已，完全是在正常的物理过程之内发生的。那位人人皆知的跳伞运动员跳出机舱之后，降落伞出了毛病，于是便直坠地面，却落到一堆干草上。这完全是他有运气。在这件事上，似乎没有上帝直接的干预。

有些人喜欢给那些罕见的巧合和大难不死的事件赋予神的意义。这种做法无非是给那些简单而罕见的自然事件以一种有神论的解释而已。但是，不管那幸运的人自己多么相信“诸神在对他微笑”，人们仍是难以用这类幸运事件客观地证实上帝的存在。一个人假如在足球赌博中发了一笔财，他自己或许也会想到，某个其他的人完全根据赌博规则也会发一笔财。那些声称获得上帝助祐的士兵在战场上杀死了敌方的士兵，不过，当敌方的士兵要取他们的命时，他们也可能要问自己上帝在哪里。

信徒：我认为，奇迹是上帝存在的最好的证明。

怀疑论者：对人们的心目中的奇迹是什么，我没有把握。

信徒：唔，就是非同寻常，不可预测的东西嘛。

怀疑论者：天上掉下来一个大陨星，地上的火山喷发，这不都非同寻常，不可预测吗？你不是说这些事也是奇迹吧？

信徒：当然，我没说这话。这类现象是自然的事件。奇迹则是超自然的。

怀疑论者：你说超自然是什么意思？是不是奇迹的另一种说法？（于是查阅牛津辞典。）辞典上这么说：“超自然的。在通常的因果作用之外的。”嗯，这全看你认为“通常的”是什么意思了。

信徒：我觉得“通常的”意思是熟悉的，或人人皆知的。

怀疑论者：我们的祖先会把发电机或无线电收音机之类看成是奇迹，因为他们不熟悉电磁学。

信徒：你说得对，他们很可能把这些装置看成是奇迹，但他们错了，因为我们知道发电机或收音机是按照自然规律工作的。但真正的超自然事件则是在任何已知或未知的自然规律中都找不到原因的事件。

怀疑论者：这难道不是个无用的定义吗？你怎么知道哪些规律可能是未知的？很可能有些完全出入意料的奇特规律，不凑巧没让我们给碰上罢了。假设你看见空中漂着一块石头，你会不会把这看成是个奇迹？

信徒：这得看……我得弄清楚这是不是幻觉或欺骗。

怀疑论者：可是，很可能有些自然过程能够产生超级幻觉，我们大家谁都想不到的幻觉。

信徒：照你这么一说，很可能我们的一切感觉也是幻觉，我们因此干脆就不用讨论什么了。是不是？

怀疑论者：好吧，咱们换一条路子讨论吧。无论如何，你仍是拿不准是不是有某种磁力或引力作用在使那块石头漂浮着。

信徒：相信上帝比相信奇异的磁现象要容易。这完全是个可信度的问题。

怀疑论者：噢！那么说，你认为奇迹实际上是“上帝造成的东

西”了。

信徒：我就是这么认为！不过，上帝有时也可能用人作中介。

怀疑论者：你要是这么说的话，就不能把奇迹说成是上帝存在的证据了。否则，你的话就成了循环论证了：“奇迹证明制造奇迹者的存在。”正如你自己也承认的，归根结底的问题是信仰。奇迹若想有什么意义，你就得先相信上帝。显然，奇迹事件本身不能证明上帝存在，因为它们可能是反常的自然事件。

信徒：我承认，从奇迹角度看，漂浮的石头是可疑的，但是，你得想想有些很有名的奇迹，例如，耶稣用几条鱼几块饼喂饱了几千人。你万不能说有哪种自然规律能使鱼和饼加倍吧。

怀疑论者：这样的故事都是几千年前由一帮迷信的狂热者写下来的，他们就是一心想促使人相信他们那一派的宗教。你能有什么可能的理由相信这样的故事呢？

信徒：你太玩世不恭了。孤立地看来，耶稣用一点饼和鱼喂饱几千人的故事算不得什么。但你得结合整个圣经来看待这个故事。这奇迹在圣经里不是唯一的。

怀疑论者：那就请给我讲讲另一个奇迹吧。

信徒：耶稣在水上行走。

怀疑论者：又是漂浮！我记得你把这一类的奇迹当作“可疑的”给打发了。

信徒：石头漂浮是可疑的，耶稣漂浮不是可疑的。

怀疑论者：为什么不是可疑的？

信徒：因为耶稣是上帝的儿子，因而，他具有超自然的能力。

怀疑论者：可你这么说，就是又把问题给推回来了。我不相信耶稣具有超自然的能力。假如他确曾在水上行走，我倒是认为这是个反常的自然事件。不过，我怎么说也不相信这样的事。我为什么要相信呢？

信徒：圣经对成千上万的人来说，一直是启示的源泉。你可小视不得。

怀疑论者：卡尔·马克思的著作也是一样。我也不相信他对奇迹的任何叙述。

信徒：你可以拒绝相信圣经的话，但甚至在近些年里成千上万的人也说他们经历了奇迹，这些人的话你是否定不了的。

怀疑论者：人什么话都说，说跟天外人见面，说远距传物，还说有千里眼。只有傻瓜或是疯子才听信这些乌七八糟的玩意儿。

信徒：我承认，很多人胡说八道，但信仰治病的证据可是很充足的。想想鲁尔底斯吧。

怀疑论者：那些被治好的人得的病都是心理原因造成的病！让我引用你的话吧："这完全是个可信度的问题。"你这话说得对。相信一些反常的医疗事件总比求助于一个神明怕要容易一些吧？

信徒：你不可能把所有的奇迹都说成是心理病。心理原因造成的病是什么意思？不就是说这样的病"从医学上讲是不可解释的"吗？假如靠信仰治好病的那些例子不过是些反常的自然事件，为什么还会有那么多的人相信？

怀疑论者：这都是魔法时代的残余。在科学兴起之前，或者说，在世界大宗教兴起之前，原始人认为，世上发生的一切，都是魔法造成的，是某个小神或小魔鬼的行为。当科学解释了越来越多的东西时候，宗教就开始走向一神论的观念，于是，魔法的解释就不行了。但

魔法思想的残余依旧存在。

信徒：你这不是说去鲁尔底斯朝圣的人都是敬拜魔鬼吧？

怀疑论者：我没有明确地这么说。但是，这些人相信信仰能治病跟非洲土人相信巫医，或跟有些人相信可以跟鬼魂往来相比，没有多大的不同。这类迷信都是魔法时代之后的返祖现象。世界上的各大宗教把这些迷信的返祖现象制度化了。谈论奇迹不过是不那么邪乎地兜售魔法罢了。

信徒：世上存在着善的神和恶的神。它们以很多方式表现自己。

怀疑论者：你认为邪恶的超自然事件也是上帝存在的证据吗？上帝也使用恶的能力吗？

信徒：善与恶之间的关系是一个微妙的神学问题。对于你所提出的问题，现在众说纷纭。人的邪恶，无论最终起源于什么，可以成为恶的发泄渠道。

怀疑论者：这么说，假如存在所谓的玄妙的能力，你认为这些东西不一定是上帝的了？

信徒：不一定是。干脆说，不是上帝的。

怀疑论者：这么说，就至少有两种超自然的事件。一种源自上帝，就是你所说的奇迹。再一种就是那些让人难受的超自然事件，我们将它们称作妖术。它们的来源尚不明确。除此之外，我想还有些不好也不坏的超自然事件，比如意念、致动、先知先觉之类。这些东西在我看来很复杂。我倒是愿意认为，这些东西都是原始的幻觉，是魔法时代的遗迹，多神论的残余。由幻觉而引起的迷信有多种多样，你对奇迹的信仰则是其中应受尊重的一种。一个具有你所说的威严和能力的上帝，是不该跟这类迷信的东西搅在一起的。

信徒：在我看来，设想有超自然的能力存在，并且设想这些能力可以用多种多样的方式加以操纵以达到善或恶的目的，这丝毫没有不近情理的地方。信仰治病是善的。

怀疑论者：而且也为上帝的存在提供了证据？

信徒：我认为是这样。

怀疑论者：那些没治好的病例又该怎么讲呢？有些不幸的人对信仰疗法没有反应，难道是上帝不眷顾他们吗？要不然，就是上帝的能力有时也失灵？

信徒：上帝的行为方式是神秘莫测的，但他的能力是绝对的。

怀疑论者：你这不过是说你不知道上帝如何行为罢了。假如上帝的能力是绝对的，他还需要奇迹干什么？

信徒：我不知道。

怀疑论者：全能的上帝主宰着整个宇宙，他可以使任何事情发生，用不着什么奇迹。假如他不想让某人死于癌症，他完全可以不让这人得癌症。实际上，我倒是想这么看：奇迹的发生证明上帝失去了对世界的控制，于是赶紧手忙脚乱地采取补救措施。你说，上帝制造的一切奇迹是什么目的呢？

信徒：上帝通过奇迹显示他神奇的能力。

怀疑论者：但他为什么这么隐晦呢？他为什么不干脆在天上清清楚楚地写一个宣言，或让月亮变色，或做些完全不可能有异议的事呢？他要是挡开一场巨大的自然灾难，或者止住肆虐的流行病岂不更好？在鲁尔底斯的奇迹中的那几个人，不管他们的病被治得多么成功，仍是跟人类的巨大苦难不成比例的。我再说一遍，你所说的奇迹在我看来不该跟全能的神搅在一起。使重物悬浮，使鱼加倍增多之类

的东西很像是戏法。这都是人类未成年时的想象的产物吧？

信徒：上帝很可能一直在为人类挡开灾难。

怀疑论者：你这是什么话？谁都可以这么说。我可以说，我每天早上念咒，于是阻止了世界大战。我而且还可以用世界大战至目前确实没有爆发这一事实，证明我的咒语有效。实际上，还真有一信仰不明飞行物体的人说这样的话。

信徒：基督徒认为，上帝一直掌握着世界，因而从某种意义上讲，世上的一切都是奇迹。有人想把自然的和超自然的东西区分开来，实际上不过是胡扯。

怀疑论者：现在，你在改变立场。你这似乎是在说上帝就是大自然。

信徒：我是说，上帝是自然界中万物的原因，尽管从世俗的意义上讲不一定是这样。上帝并不是把一切开动起来，然后就撒手不管了。上帝是在世界之外的，是超越自然的一切定律的。他维持着世界的存在。

怀疑论者：在我看来，我们这似乎是在争论一个语义问题。大自然有一套优美的定律，宇宙就照着这些定律划出的路线一路演化过来。你所说的“上帝维持着世界的存在”，只不过是从有神论的角度说了同一件事罢了。上帝仅仅是一种说法而已，不是吗？说上帝维持着世界的存在跟说宇宙持续存在有什么不一样呢？

信徒：你不能仅举出宇宙存在这一明显的事实就算完成任务了。宇宙还得有个解释。我认为上帝就是宇宙的解释。宇宙存在就是一个奇迹，为了维持宇宙的存在，无时无刻不需要上帝的能力。在大多数情况下，上帝用一种有秩序的方式来维持宇宙，也就是你所说的，用物理定律来维持。但上帝有时也脱离这秩序，制造些惊人的场面来向

人类发警告或信号，或向信上帝的人提供帮助。例如，他曾为希伯来人分开红海的海水，使他们能安然走过红海，甩开法老的追兵。

怀疑论者：我觉得难以理解的是，你怎么会认为这一个超自然的奇迹的制造者就是那位宇宙的创造者，就是回应祷告、创造物理定律、审判众生的那一位呢？为什么这些超自然的奇迹的创造者不能是不同的个体呢？世上的奇迹这么多，各种不同而且还相互冲突的宗教显然都可以由这些奇迹得到证据，我倒是想过，相信奇迹的人应该承认有一大群超自然的存在物相互竞争才对。

信徒；有一个上帝要比有很多上帝来得简单。

怀疑论者：我仍是不明白，那些所谓的奇迹，不管有多么不同寻常，怎么能被看成是上帝存在的证据。在我看来，你似乎是也相信神话中的那位主宰众生的女神。不过，我们都有这样的本能，只是你把“幸运女神”代换成一个实有存在物，把她称作上帝罢了。你怎么能真相信那些“奇迹”呢？

信徒：我没有发现上帝有什么不可思议之处。他是万能的创制者，他操纵着一切物体。与上帝所创造、所维持的宇宙这一奇迹相比，上帝分开红海海水又有什么好奇怪的呢？

怀疑论者：但你说这话时，立论的基础是设定上帝存在。假如存在着一个你所说的那样的上帝——无限、全能、仁慈、全知——那么，我也会认为分开红海海水对这样的上帝来说是小事一桩。但问题是我们怎么知道他确实存在？

信徒：这完全是个信仰问题。

怀疑论者：完全正确！

我希望，上面的这一大段不了了之的对话，能够说明科学和宗教在超自然问题上的主要分歧。信教的人相信上帝的全能，并且在日常

生活中到处都能看见上帝的作为，因而在他们看来，奇迹没有什么不合理之处，因为奇迹是上帝在世界上的作为的另一个方面。科学则与信教的人截然相反，他们认为世界是根据自然定律运转的，奇迹是“反常”，是一种病态，破坏了大自然的优美。奇迹是大多数科学家愿意撇开的东西。

当然，奇迹的证据具有高度的争议性。假如只是根据现存的证言承认有奇迹，那么，我们就没有什么好理由拒绝承认许多其他的所谓奇迹（不明来源的飞行物，幽灵、意念弯物、测心等），因为这些奇异的事件似乎也同样有人证实。但即使一个科学家相信了奇迹，我们也很难在奇迹和异常现象之间划一个明确的分界线。

现在，在人们中间正在兴起一股对意念弯物和超感官知觉之类异常现象的巨大兴趣。异常现象的研究者很少使其研究对象带上神学的涵义。即使是信念治病之类的异常现象也被看作是“非上帝所为的奇迹”。很多人带着原始的狂热信仰对异常现象进行研究，而这些信仰却使宗教的名声受到损害。一家有名的报纸的星期日副刊曾把耶稣基督和尤里·盖勒[3]等量齐观。不幸的是，很多得到报道的奇迹都有变戏法的味道。库帕第诺的圣约瑟夫据说就很让他的信教的兄弟们为难，因为他在做礼拜时，动不动就浮入空中。于是，信主的众兄弟们为了能好好作弥撒，到时候就把他关在他的小室里！

有趣的是，所谓超自然的宗教事件的很多象征现已重新出现在现代人对不明飞行物的狂热崇拜之中。例如，有些自称身患痼疾的人说，他们跟不明飞行物中的人接触之后，病立刻就好了。偶尔还有人只是看见了不明飞行物，病就好了。

重物悬浮空中也引人注目。有人告诉我们，那些转瞬即逝的飞碟之所以能在天上无声地高速飞行，靠的不是笨拙的火箭，也不是大马力发动机。飞碟飞行，靠的是中和地球的引力得来的动力。有时，不明飞行物的驾驶者也在地平面上飞来飞去而不下坠。

显然，天外异物，重物悬浮和信念治病之类是深深地根植于人类的精神之中的。在魔法时代，这些东西是堂而皇之的，公开的。随着有组织的宗教的兴起，这些东西变得不是那么赤裸裸了。但其强烈的原始成分从没有往下压得很深。现在，随着有组织的宗教的衰落，这些东西又借着科技的伪装重新泛起。信仰这些东西的人也满口宇宙飞行器、伪科学、神秘的力场、精神作用于物质之类的时髦名词，其实，他们这一套是原始迷信和空间时代物理学的大杂烩。

奇迹历来就是宗教出风头的一个方面，过去因为跟其他那些所谓反常现象并列在一起而遭诟病。这是由于很多如妖术之类的反常的东西似乎令人极其厌恶。相信奇迹的人若想要说服怀疑论者，便面临着双重的难题。首先，他得说服怀疑论者相信，那些可称作奇迹的反常现象确实发生了，而做到这一点是极其不容易的，因为他所得到的大部分证明都是可疑的。然后，他得让怀疑论者相信奇迹是跟上帝有直接关系的。这就是说，他要么得承认一切超自然的事件（甚至那些令人不快的事件）都是上帝的作为，要么得设法在上帝所行的奇迹和其他的反常事件之间划出一道明确的界线。在这超感官知觉家喻户晓的时代，大部分相信有奇迹的人宁愿把宝押在思维的能力上，而不押在上帝的能力上。

# 第十五章　宇宙的终结

尘世繁华转眼即逝。

假如宇宙是上帝设计的，它就一定有一个目的。假如这目的永远也达不到，上帝就会设计失败。假如目的达到了，宇宙就没有必要继续存在下去。至少就我们所知，宇宙将有一个终结。

在宇宙死灭的时间和方式问题上，各种宗教分歧很大。有的宗教警告人们说，宇宙的灭亡就在眼前，到时将有启示录上所讲的大毁灭，罪人将受到严厉的审判。其他的宗教则教导人们，天国即将来临，我们现在所看到的恶劣而无常的世界将被取代。有些东方宗教则倾向于认为，世界处于轮回之中，这个世界的终结便是预示着另一个世界的再生。

现代科学对宇宙的终结有何看法呢？

在第二章里，我们曾说过，热力学第二定律在无情地削弱宇宙的组织，使宇宙进入混沌。我们在宇宙的每一个角落都能看到，熵在不可逆转地增大，庞大的宇宙秩序在慢慢地然而也是确实地被消耗。宇宙似乎注定要继续衰败下去，走向一种热力学平衡状态，即达到最大的无序状态，然后就再也不会发生什么令人感兴趣的事了。物理学家们把这种令人沮丧的前景称作“热死”。100多年来，人们一直在谈论热死。

热力学第二定律对整个物理学来说是基本的定律，因而没有多少物理学家对其有效性提出怀疑。我们在第九章里看到，第二定律使世

界带上了时间不对称性，使得过去和将来有了分别。违反热力学第二定律就等于将时间倒转。

然而，第二定律并没有讲述那些把宇宙驱向最大的无序的终结状态的灾难具有什么性质。在过去的30年里，随着现代天文学的飞速发展，已有可能从一些细节上详细说明那些最终将要毁灭宇宙的复杂组织的事件了。而且，我们也有可能详细说明我们周围的世界的活动了。

就我们所处的宇宙区域而言，地球的命运是与太阳的命运紧密相联的。地球上的生命以阳光为生，对太阳现状的任何大的破坏都会造成灾难。太阳很可能有什么变故，一有变故，地球就住不了人了。对太阳恒定的热输出的任何改变都能打破地球脆弱的气候平衡，使我们进入灾难性的冰期。与太阳风（即来自太阳表面的稳定的粒子流）相关的太阳系磁场的变化也可能带来同样的灾难性后果。我们地球附近若有哪个恒星发生了爆炸，便可能使我们陷入致命的辐射之中，某个黑洞穿过太阳系也可能使众行星的运行轨道发生变化。

但是，假如地球能逃脱所有这一切令人不快的可能性，它显然也不能永远维持现状，“永世长存”。太阳所辐射出来的大量的能量，得要用核燃料来补充，而太阳的燃料储备最终会用光的。天体物理学家估计，太阳耗尽燃料还得要40～50亿年。宇宙现在的年龄是180亿年，而太阳的年龄已在45亿年，正处在鼎盛的中年。

随着燃料的减少，太阳会膨大起来，变成天文学家所说的一种红巨星。太阳的内核在艰难地维持能量生产的同时，会收缩又收缩，直至发生了量子效应才使它稳定下来。在这个阶段，太阳可能膨胀得很大，以致太阳附近的行星会被它吞没，地球的大气也被剥离，岩石熔化，甚至汽化。以后，太阳会开始进入一个飘忽不定的生涯，现在十分充沛的氢燃料的核反应到时将会为效率不那么高的氦燃料核反应所取代，氦燃料烧完之后，将会有更重的、再更重的元素燃烧。这些我

们已在第十三章里说过。

当最后所有的燃料耗尽时，太阳就是由铁一类的较重的元素构成的了。那时，核聚变不会再释放能量。铁是原子核的最稳定的形态。根据热力学第二定律，一切系统都寻求达到其最稳定的状态。在这一阶段，太阳的中心温度会稳步上升到接近10亿度。燃料耗尽之后，太阳内部外向的压力将减弱下去，引力将占据支配地位。一蹶不振的太阳将开始在自己的重压下收缩，使其内部的物质受到剧烈的挤压，其物质的密度将高达每立方厘米100万克。燃尽后收缩的太阳将会变得跟地球一样的大小，然后在多少亿年里半死不活地慢慢暗淡下去，最后冷却成为一个黑矮星。

在我们的星系以及所有其他的星系之中，到处都会有这同一套模式的重复：恒星先是不稳定，然后是燃料枯竭、恒星崩塌。所有的恒星会一个接一个地烧完自己的核燃料，最后再也支撑不住自身的重量，终于让无情的引力摧垮。

有些恒星（如超新星）死得颇为壮观。超新星内核先发生惊心动魄的塌缩，将超新星自身炸成碎片，释放出巨大的能量。这些突如其来的超新星爆炸之后，其中较轻的将变成一些弥漫的碎片，环绕在一块高度压缩的物质周围。在这高度压缩的物质之中，质量相当于太阳的物质能被压成一个直径只有几英里的圆球。这样的物质重量巨大无比，一调羹这样的物质就比地球上所有大陆的重量加在一起还重。如此巨大的重力连原子也承受不了。于是，原子被迫向里崩溃，变成了一些纯粹的中子。天文学家们熟悉中子星，因为他们在过去的超新星的爆炸碎片中时常发现中子星。

较重的死星在面对引力巨大无比的情况下，甚至不能够以变成中子星的方式来稳定自身。它们会加速收缩，最后变成黑洞。

宇宙学家爱德华·哈里森描述了宇宙缓慢衰亡的过程，他的语言很

是形象生动：

所有的恒星将会开始像即将燃尽的蜡烛一样暗淡下去，然后一个一个地熄灭。在空间的深处，那些宏伟的天体城邦，那些星系，将满载着多少时代的历史记录缓缓地死去。在成百亿年的岁月里，宇宙会越来越黑暗。偶尔会有几点光亮划破宇宙的夜幕，阵阵短促的天体活动使注定要变为星系坟场的宇宙得以苟延残喘。①

物理系统在寻求最高熵状态的过程中探索了一些古怪的途径。随着我们所在的星系当中的恒星无情地燃尽，我们星系的组织也会开始瓦解。太阳之类的恒星若要燃尽，得需要几十亿年的时间，而在这段时间里，新的恒星将会不断地从星际气体中产生出来。较小的恒星走向死亡所花的时间，可能要几千倍于太阳死亡所花的时间，然而，锁闭在恒星之中的有序的能量最终还是以辐射的形式杂乱无章地散布到宇宙中去，我们所在的星系将会暗淡、冷却下去。其他的星系也会遭到类似的命运。

那些业已死亡的恒星仍会有大量的活动，但其活动的时间尺度大大增加了。恒星燃尽之后剩余的残骸在星系中漂游，时时会有碰撞发生。黑洞会吞没任何恒星以及它所遇到的其他物质。而且，假如像某些天文学家所认为的那样，我们的星系的中心有一个大黑洞，那么，这个黑洞将会越变越大。黑洞放射出来的引力辐射会使恒星的轨道慢慢地破坏，因为引力辐射是空间的波纹，这些波纹会使一切大质量物体的轨道能量衰竭下去。在极长的时间里，恒星残骸会越来越靠近星系中心，最后被那永不知饱的黑洞吞没。有些死星却能逃离这一厄运，这是因为它们幸运地撞上了其他的恒星，这些恒星将它们撞出星系，成为星系之间茫茫无涯的空间中的孤独的漫游者。

对这样的死星来说，以及对所有逃脱死于黑洞这一厄运的气体和尘埃来说，这逃脱只是暂时的。假如大统一论是正确的，那么，这些宇宙流浪者的核物质就是不稳定的，大约$10^{32}$年之后，它们的核物质

就会蒸发干净。中子和质子变成正电子、电子，正电子和电子又进而相互湮灭，并湮没任何其他再生的电子。一切固体物质就这样分解了。这种大屠杀的最终结果是什么，要视宇宙实际膨胀速度而定。假如真是像有人估计的那样膨胀得较快，那么，快速膨胀的宇宙就会把电子和正电子扯开，使它们不能相撞，因而也就不会发生完全的湮灭，宇宙当中总是剩下一些粒子。那些湮灭的粒子则产生伽马辐射，伽马辐射本身也随着宇宙的膨胀而缓慢地减弱。除此之外，还有大爆炸残留下来的中微子和热辐射。所有这些东两都会逐渐冷却下来，温度跌向绝对零度，但它们彼此间的降温速度有所不同。物质（电子和正电子）冷却得比辐射快。因此，尽管物质和辐射都走向绝对零度，它们的温差也逐渐减小，但是，它们之间总有一定的温差，而这温差原则上讲是能用作自由能量（负熵）的来源的。因而，尽管这高度衰竭的宇宙的熵已接近其最大值，但从未达到其最大值，所以，在这一限度之内，热死永远也不会发生。

假如宇宙膨胀得较慢，电子和正电子的湮灭就会更容易发生。然而，电子和正电子相互毁灭并非是单纯的偶然碰撞的结果。电磁力使电子被吸引到正电子那里，使它们能够形成一些被称作电子对的“原子”。计算显示，在缓慢膨胀的宇宙中，大部分粒子在$10^{71}$年之后都会成为电子对，但电子对这种原子实在是稀奇古怪，竟有千万亿光年那么大！这些粒子缓慢地环绕着彼此旋转，它们运动1厘米要花100万年的时间。电子对是不稳定的，它们的巨大轨道会因为它们发射的很低能的光子而非常缓慢地损坏下去。$10^{116}$年之后，大部分电子对将会崩溃，所有的粒子将发生接触，于是立即会发生湮灭。在电子对的轨道崩坏期间，每个电子对“原子”将放射出不少于$10^{22}$个光子，使熵大大地增大。

黑洞也并非一直静止不动。在第十三章里简要讨论过的量子效应表明，黑洞严格地说并不黑，而是借着热辐射一直在发着幽暗的光。一个质量跟太阳一样大的黑洞，其温度低得可怜，只有热力学温度的

零上一百亿分之一度，而超级黑洞的温度更低。只要宇宙的背景温度高于它们的温度，黑洞就会通过吸热而非常缓慢地持续增大。当黑洞与其他物体或其他黑洞相撞时，仍然会出现一些活动，随着黑洞自旋的消耗，黑洞的旋转会逐渐慢下来。但是，当空间的温度最后降到黑洞的温度以下时，便会出现最为激烈的变化。

比其周围环境热的黑洞会损失热量，因而也就损失了能量。能量的损失将会使黑洞收缩。而黑洞收缩又会使温度稍微增高，使能量辐射加速。于是，黑洞便开始滑向遏制不住的蒸发过程。在极长的时间里，黑洞的收缩速度一直向上攀升，最后，大约经过$10^{108}$年之后，那些起初比很多星系加在一起还重的黑洞将会缩得一干二净，无影无踪。

现在，人们都还不知道黑洞会如何死亡，但黑洞似乎有可能缩得极小，变得很热，于是开始创造物质。但黑洞在这一阶段只能存活几十亿年。最后，黑洞很可能爆炸，变成一些伽马射线，一点往昔存在的残迹都不留。

上述这些研究显示，我们现在所看到的这个充满活动、富丽堂皇的宇宙将会有一个惨淡的下场。尽管宇宙灭亡所需时间长得令人无法想象（请记住，$10^{100}$是1后面有100个零），然而，看来没有什么疑问的是，我们目前所观察到的一切结构注定要消失，只留下那黑暗、寒冷、近乎空无一物的空间，其中的物质密度越来越低，只有几个零星的中微子和光子，以及很少的其他东西。很多科学家觉得这个场面很令人沮丧。

然而，宇宙还有另一种下场。之所以得出上述让人沮丧的结论，是因为假设宇宙会一直膨胀下去。这个假设并不确实。人们已知，宇宙的膨胀速度一直在稳步地降低，因为引力在抑制着星系分离。有些天文学家认为，宇宙的膨胀有一天会停下来。实际情况是否真是这样，要由宇宙引力的大小来定，而宇宙引力的大小又要由物质的密度

来定。因为宇宙物质包括看不见的物质（如中微子和黑洞）以及看不见的能量（如引力波），所以，要想估计总体的物质密度几乎是不可能的。

假如宇宙的膨胀真的停下来了，宇宙也不会保持静止。宇宙会开始收缩，其收缩运动是其膨胀阶段的时间映像。开始的时候，收缩是缓慢的，但几百亿年之后，收缩的速度就会加快。各星系现在是彼此相离而去，那时就会开始相互靠近，而且靠近的速度越来越快。宇宙的收缩阶段会酿成大灾难。

当宇宙缩至现有大小的1%时，收缩的效应会使温度高到水的沸点，地球（假若太阳在做垂死挣扎时地球有幸不死的话）将会变得无法居住了。那时，观察者也不能分辨一个一个的星系了，因为随着星系间的空间靠到一起，星系也就开始彼此融合了。宇宙若进一步收缩，温度还会进一步提高，那时，天空本身也会变成火炉，在这白热化的空间之中，恒星也会开始沸腾，然后爆炸。

这时，各个事件的进程加快了步伐。所有的结构都会消散，它们的原子也分崩离析。在短短几千年的时间里，原子核本身也会在迅速增高的温度中被击成碎片。这时，事件的时间尺度将小得不可思议。宇宙先是在几分钟的时间里收缩一些，然后又在几秒钟，又在几微秒钟里收缩很多。累加的引力将宇宙的收缩变成了失控的向内爆炸。这，就是所谓的“大崩塌”。

这些可怕的事件引发了诗人诺曼·尼科尔森的灵感，使他写出了这样的诗句：

假如宇宙

倒转并现出

它的本相；

假如可见的光

向内流逝，自天空暴风雪一样的降下

众多的星系。

黑夜的透镜就会烧得

比聚焦的太阳还亮，

人就会失明，

因为他眼中是一片白热的黑暗。[2]

这时的宇宙离死亡就只有几微秒了。大崩塌就像倒转的大爆炸。核粒子分裂成夸克，在短暂的时间里，所有的亚核粒子都被创造了出来，但是，整个宇宙一下子就在一瞬间缩得比一个原子还小，于是，时空本身也分解了。

很多物理学家认为，大崩塌就代表了物质宇宙的终结。他们认为，宇宙（一切空间、时间、物质）诞生于大爆炸，他们也同样认为宇宙消失于大崩塌。消失于大崩塌就是完全的湮灭，什么也留不下，地点没有了，时刻没有了，什么东西都没有了。当存在的一切都死于引力的无限的毁灭力量时，就有了最终的“奇性”，然后就再也没有什么了。引力曾是宇宙的接生婆，引力也是宇宙的送葬者。

然而，并非所有的科学家都愿意相信宇宙真会这么轰轰烈烈地死去。有些科学家认为，一些未知的物质力将会在宇宙收缩到某一极高的密度时使大崩塌停下来，使宇宙“反弹”回去，进入另一轮膨胀和收缩的过程，如此这般循环往复以至无穷。在第十二章里我们已经提到过这种振荡的宇宙。只有在超高能物理上作进一步的研究，才有可能解决这个问题。

尽管科学为宇宙设想的下场多种多样，五花八门，但所有的下场都涉及我们今天所知的宇宙的死亡。在这一点上，科学所设想的宇宙下场与最具有宗教性的末世学是一致的。然而，宇宙死亡过程的时间尺度大得不可想象，所以不可能把宇宙的死亡同人类的活动联系起来。假如有哪种有意识的生物生存在如此遥远的将来，以至现今的时代在它们看来无法与宇宙创生分别开来，那么，这种生物不会是人。它们到底是什么，还是让千万亿年的演化和技术进步去管吧。

首先，“人工”智能的发展就很可能意味着人类将把自己在智力方面的最高地位转让给会思想的机器。实际上，这种情况在一有限的意义上已经发生了。将来的技术革新的时间还很长很长，机器似乎没有什么理由不能够做到，并且是更好地做到人脑所能做出的任何事情。因为这种智能机器没有尺寸的限制，所以不难想象将来会出现巨大的人工超级脑，这种超级脑的智力是我们现在完全理解不了的。而且，电子装置能够直接在彼此间传输信息，这也为电脑的综合开辟了道路。可以想象，有一天宇宙中会出现一个复杂的无线电通讯网络，把数不清的散在各处的超级脑连接成为一个单一的极大脑。

基因控制方面的进展能使人们对思想机的概念发生新的变化。到目前为止，生物智能的发展一直是听命于自然的进化力量。但是，随着我们对决定着我们的肉体和精神特点的分子结构取得了控制权，就有可能修改现存的生物，甚至有可能发明新的生物。杂交和突变诱发已在有限的规模上做到了这一点。现在似乎没有什么基本的理由怀疑那一天的到来。到那时，可以通过分子工程按订单来“培植”大脑。那时，自然智能和人工智能的区别将会消失。这些人类创造的高级脑既可被看作是基因受控的生物体，也可被看作是由有机硬件而不是由固态硬件构成的高级计算机。现在甚至可以设想，人造生物脑和固态硬件能够共生，即有机脑可“插入”固态电路。或者，未来的超级电路片可以像一种“成套的放大器”一样插入大脑。同时或许有可能在较为常规的思想机里用有机的零件来代替某些半导体晶体。当然，现在没有

谁说上述的可能性在可见的将来有任何可行性，但是，难道我们不可以真的相信，再经过100万年、10亿年、1万亿年的科学研究，这些可能性会成为现实吗？要记住，科学的年龄才仅仅几百年啊。

一个与宇宙遥远的将来以及宇宙当中的居住者相关的问题是，智能生物对自然界所能施行的控制是否有一个限度。我们所看到的宇宙是由各种巨大的宇宙力形成的。宇宙力包括强大的核相互作用和引力的远程作用。但是，我们也看到一些初步的人造环境：河流被改道、被拦截，森林被培植起来又被毁掉，沙漠被驯顺，山头被削平。在地球的表面，没有任何人类活动留下足迹的区域不多了。随着技术与科学的发展，我们可以期望我们的后代对更大的、更复杂的物理系统取得控制权。弗里曼·戴森曾经想到，天外有些技术先进的生物群落大大改变了它们的太阳系的结构，它们为自己的太阳制造了一个球形的物质壳，以最大限度地截留并利用太阳的能量输出。给行星剥去衣服所需的那一等级的技术，可能看上去永远是一个幻想，但进行这浩大的工程所必需的东西首先是时间、金钱和资源，而不是技术。

因而，我们如今所面对的就是令人迷惑的前景。在一个拥有几乎无限的时间进行技术革新的宇宙中，我们能有把握地排除任何与物理定律一致的东西吗？在过去的短短几千年里，人类在技术上就取得了飞跃的进步，从开始时只能制造几厘米大小的工具，到现在能进行多少英里的浩大工程（桥梁、隧道、城市）。假如这种技术进步的趋势持续下去的话，那么，即使以后的进步速度大大减慢，也终究会有一天，整个地球，然后是太阳系，最后是所有的恒星都将被“技术处理”。我们的星系可能被改造得面目一新，一些恒星被从原来的轨道上挪走，可以用气云制造新的恒星，也可以用人工制造的不稳定来处理掉一些恒星。人们还可以随意制造或控制黑洞，使黑洞成为能源，或成为宇宙社会的废物处理装置。

假如星系可以这么人工处理，为什么整个宇宙不可以？

这样的推理或许会被斥为荒唐，但它们的确提出了一个重大的哲学问题。假如自然的东西和人工的东西之间、盲目的力量与智能的控制之间有什么区别的话，这区别是什么呢？这正是从一个新角度来看待自由意志和决定论之争。

当某一系统被置于人工智能控制之下时，该系统仍然是与物理定律一致的。目前没有任何证据显示，较大的人工建构违反任何物理原理（除了在精神-肉体相互作用的层面之外，但在这一层面上是否真的违反了物理学原理，人们仍是在争论）。的确，一个铁路网或一个核电站不会自发地出现，但是，铁路网和核电站的建造仍是在大自然定律的框架之内进行的。人们进行这些建设所得到的秩序被建造过程中产生的熵所抵消。

我们在第六章里讨论过，大脑的工作要分两个层面进行描述。一个层面是硬件层面，要用物理规律来描述。另一个层面是软件层面，是与硬件层面等价的、一致的。这软件层面要用思想、感觉、决定之类的东西来描述。同样，说某一系统被“技术处理了”，并不是说物理定律的权威性被否定了，这只是用软件语言来描述该系统的运行而已。那么，宇宙既按明确的物理定律运行，同时又在智能控制之中，这二者就没有什么矛盾的了。

这是一个发人深思的结论。那些乞灵于上帝来解释宇宙的组织的人，通常脑子里想的是一个超自然的行为者以违反自然定律的方式作用于世界。其实，我们在这个宇宙中遇到的很多东西完全有可能是纯自然的、在物理定律之内的智能控制的产物。例如，我们的星系就可能是一个强大的智能者制造的。这位智能者精心置放引力物体，控制天体爆炸，使用所有其他各样空间时代天体工程师的器具，重新安排了原初气体，从而制造出我们的星系。但是，这个超级智能是上帝吗？

这个问题非同小可。上帝通常被认为是整个宇宙（包括时间与空

间）的创造者，而非仅仅是个星系建筑师。显然，必须在物质宇宙之内只利用早已存在的定律进行工作的那一位，不能被看作是宇宙的创造者。但是，假如这位超级天体工程师能将其能力施展到所有的星系上呢？我们可以设想，他能够利用引力来使空间和时间弯曲。

然而，假如这超级天体工程师不能真正地创造或毁灭时间和空间，他就不是上帝。可是，新物理学在这里提出了一个有趣的意见。假如有足够的能量和资源，那么，人类也能够积累起足够的引力物质，制造出一个黑洞。在这黑洞的中心，在所谓的奇点处，时间和空间被毁灭了。于是，连我们都能毁灭时空。

创造时空要比毁灭时空难。但是，我们真能有把握地确信，创造时空真是不可能的，真是完全为物理定律所不容的吗？我们不敢说这话。实际上，在第三章里，我们说过近来有一些理论认为，在大爆炸中空间的“泡”被创造出来。而且，假如宇宙创生于大爆炸这种流行理论不对，空间和时间真是永恒的呢？假如空间和时间一直存在，谈论宇宙在时间当中被创造的事就没有任何意义。上帝在宇宙之内的作用就仅限于形成并组织物质了。而形成并组织物质这种事，上帝或许完全用自然的手段就办得到（在这里，我们绕开了某些热力学的问题）。按照这种观点来看，上帝可以是宇宙中的永恒的、无限的、最强大的神灵。但他不是全能的，因为他不能不按大自然的定律行事。他可以是我们所看到的一切的创造者。他用先前存在的能量造出了物质，对物质进行适当的组织，创设必要的环境使生命得以发展，他做了诸如此类的很多的事。但是，他不能像基督教教义所说的那样，可以从无中创造。这样的上帝是一个自然的上帝，而不是超自然的上帝。

我们有什么证据能够证明有这样一个自然的上帝呢？证明自然的上帝存在的证据要比证明超自然的上帝的存在的证据更强还是更弱呢？

自然界里有很多未解之谜。这些谜可以通过假设有一个自然的上帝存在而很容易地得到解释。例如，星系的起源现在没有令人满意的解释。生命的起源是另一个令人头痛的谜。但是，我们可以认为，星系和生命这两种系统都是由一个超级智能者在不违反任何物理定律的情况下有意识地制造出来的。然而，这样的解释把任何现今的科学所不能理解的东西都归因于上帝，因而落入了窠臼。（这样的上帝被讥为“救驾的上帝”。）在付出了代价之后，笃信宗教的人也已明白，最危险的是指着一个现象说“这正是上帝作为的证明”，结果却发现后来的科学进步给这一现象提供了完全充足的解释。把上帝推出来对未曾解释过的东西进行一番总括性的解释，这只能引来弄虚作假，并使上帝与无知为伍。假如我们想找到上帝，那肯定只能通过我们在这世界上发现的东西来找，而不能通过我们没能发现的东西来找。

然而，在这一类的争论中，相对于超自然的上帝而言，自然的上帝更胜一筹。一个自然的上帝在物理定律的范围里创造了生命，这样的假设至少被认为是可能的，是与我们对自然界的科学知识一致的。因为人类在实验室里创造出生命都具有明显的（尽管是遥远的）可能性。

对生命（或其他任何高度有序的系统）的起源有两种解释。一种是生命来自一个超级智能者，或可能是至高的上帝施行自然的智能控制的产物。一种是生命是盲目的自组织过程（如木星大气中出现的有序的环流花纹）的终端产物。这两种解释都有难以说得通的地方。

这个问题的答案必须决定于我们在多大程度上认为精神是宇宙间的一个重要力量。大多数人都愿意相信科幻小说所描绘的那种遥远的未来，那时宇宙当中会有更多的区域被置于智能的控制之下。人们可以设想，再过多少万亿年，我们现在所能观察到的整个宇宙将被技术加工过。既然如此，设想这位对宇宙进行技术加工的超级智能者不可能先于我们存在，这又有什么困难的呢？

传统的观点认为，经过一个很长系列的变化，逐渐地增加了物质的组织程度之后，智能才会作为变化的终端产物出现。说得简单些就是，物质第一，精神在后。但是，实际情况真是非这样不可吗？精神难道就不可能是那更原始的存在物吗？

现在在科学家中间，越来越占上风的意见是，精神和生命都不必仅限于有机物质。物理学家杰拉尔德·费恩伯格和生化学家罗伯特·夏皮罗最近出了一本极具推理性而发人深思的书——《地球之外的生命》。在这本书里，他们重新探讨了地球之外的生命的各种可能性。他们认为，只要有了等离子体、电磁场能、中子星的磁区以及其他一些各种各样的稀奇的系统，就会有生命存在。现在，意识和智能都是软件概念。软件的关键只是模式，是组织，而不是其表达媒介。由此得出的一个逻辑结论是，我们可以设想，自宇宙创生以来就有一个超精神，这个超精神包含着大自然的各种基本场，是它承担下来那个艰巨任务，把无条理的大爆炸转化成为我们现在所观察到的复杂而有序的宇宙；它所做的这一切，都是在物理定律的框架之内完成的。这样的上帝不是用超自然的方式创造一切的上帝，而是一个施行指导、控制的普遍精神，这精神弥漫于整个宇宙，操作着大自然的各种定律以达到某种特定的目的。对此，我们可以这样说：大自然是它自己的技术的产品，宇宙是一个精神，是一个自观察、自组织的系统。我们人的精神可被看作是宇宙精神的海洋中一些局部的意识“孤岛”。这种看法令人想起东方的神秘主义的观念。在东方神秘主义看来，上帝是那个统摄万物的意识，当人的精神进步到适当的阶段时，就会被那统摄万物的意识所吸收，从而失去其独立的个性。

还可以往下继续推理。要记住，至少在某些物理学家看来，精神对量子因素而言具有特殊的地位。假如说，精神能够“加重量子骰子”从而改变掷骰子的结果，那么，一个宇宙的精神原则上则能引导每一个电子、每一个光子、每一个质子等等的行为，控制宇宙中所发生的一切。当我们观察微观物质，我们注意不到有这样的组织力量，

因为任何一个特定粒子的行为仍然在我们看来似乎是完全随机的。只有在大量原子的集体行为中，组织才会显现出来，而我们看不见这样的集体行为，于是便会说宇宙系统是以未知的方式自组织的。把上帝描绘成这样一种精神，或许足以让大多数信上帝的人满意。

很多早期宗教都是多神论，认为有很多神，这些神按其能力不同而各有等级。这种观点在现代人对天外智能的思考中也有表现。有些作者设想，智能和技术能力存在着高低不同的等级，在这等级之中，人的级位最低。可以设想一些生物具有很了不起的能力，以至我们没法把它们的行为同大自然本身区别开来。在这等级次序中有一个至高者，这至高者的能力和智能最高。这位至高者能够做到很多传统的上帝所能做到的事。

假如我们相信的确有一个这样的精神存在（从科学上看，尚没有任何可能的理论真正地证明这样的精神存在），那么，这精神能够不让宇宙走向灭亡吗？

假如这位至高者只能被迫在物理定律的范围之内活动（尽管量子论容许物理定律可有相当的变通），那么，就可以肯定地说，这样的精神不能阻止宇宙的死亡。热力学第二定律禁止任何人将那无情地上升的熵降下来，不管这人的技术多么高超，也不管他的知识多么到家，他就是不能让熵降低。

人们也许会这么想：那个能够在原子水平上操纵物质的神也能够通过恢复宇宙逐渐松弛的组织形式，不断地给宇宙“重新上弦”。这种想法其实是麦克斯韦在一个世纪以前探讨过的，人们通常称之为麦氏妖的悖论。我们可以考虑一下，有一个密封的箱子被一个带开闭器的薄膜一分为二。薄膜的两边都是温度和压力相同的气体。这箱子是一个系统，目前处于最大熵的热力平衡状态，没有任何可以做功的有用能量。在这个系统中，除了气体分子的随机跳跃之外，不会进一步发生什么有趣的事。

然而，假设在这箱子里有一个小妖怪能够操作开闭机关，它注意到分子的运动混沌无序，有各种各样的速度和方向，有的分子运动得快，有的运动得慢。分子的平均速度是决定气体温度的要素。平均速度不会改变，但每个分子每次撞上邻近的分子或箱壁时，其运动速度和方向都会改变。于是，小妖便采取了下面的策略。当箱子右边的快速运动的分子接近薄膜时，它就打开开闭器，让这分子穿过薄膜进入箱子的左边。反之，当箱子左边慢速运动的分子接近薄膜时，就让它进入箱子的右边。过了一段时间以后，箱子的左半部将满是快速运动的分子（平均而言），而箱子的右半部将满是运动速度较慢的分子。箱子的左半部因而要比右半部温度高。这妖怪就这样灵活地操纵个别分子，使箱子的两个部分造出温差。热力学平衡将不再占上风，于是熵也减小了。现在，可以用这温差来完成某种有用功了（例如开动热机），直至最后的能量又被消耗掉，热力学平衡又得到恢复。然后，那妖可以故伎重演，我们也就有了粗具规模的永动机了。一个无所不在的妖可以在宇宙水平上玩弄这套分选把戏，使宇宙不会坠入热死之中。

可惜的是，精心的观察证明，麦氏妖干不了开动永动机的活儿。在20世纪20年代，列奥·斯济拉得详细地研究了麦氏妖的工作情况。他发现，那妖若想把事儿干成，就必须有确切的信息，知道向它接近的分子的运动速度有多大。只有付出一定的代价才能获得这样的信息，而付出的代价就是熵的增加。例如，妖可以用强的光流来照亮近前来的分子，并用多普勒效应来测量它的速度，就像警察用雷达给汽车测速一样。但是，妖在这一步骤上花费有用的能量本来是要分选速度不同的气体分子以使熵降低的，而实际上则会使气体的熵增高。很明显，智能控制在分子水平上不能够击败热力学第二定律。

假如这些热力学的理论是正确的，那么，任何自然的作用，无论是智能的还是非智能的，都不能永久地延缓宇宙末日的到来。我们已经说过，假如宇宙持续膨胀下去，就可能永远也达不到真正的热力平

衡。然而，我们今天所看到宇宙生机勃勃的组织注定要衰落得面目全非。只有超自然的上帝才能真正地使宇宙再次振兴起来。

# 第十六章　宇宙是“免费午餐”吗？

从无中不可能创造出任何东西来。

卢克莱修

现在，我们可以把所有的研究线索聚拢到一起，为宇宙创制一套说明，以显示新物理学在多么令人惊讶的范围里解释了物质世界。我这并不是说，应当把我们要创制的说明当真（尽管物理学家们正在认真地讨论它）。然而，这套说明的确显示了现代物理学所提出的那一种思想。我们若想寻找上帝，就不能忽视这一类的思想。

在本书的前言里，我曾提出我所说的有关存在的四大基本问题：“为什么大自然的规律是现今这样的？为什么宇宙是由组成它的各件东西组成的？这些东西是如何起始的？宇宙如何获得了组织？”

为了解答这四大问题，新物理学走过了漫长的道路。我们逆着这四大问题的顺序，讨论了假如有负熵输入，起始的混沌状态如何能演进为较为有序的状态。我们也讨论了宇宙的膨胀如何能够生出负熵，因而也就不必像早些时代的科学家们那样，设定宇宙创生于具有高度组织、特殊排列的状态。宇宙现今的组织是与那个偶然起始于随机状态的宇宙不矛盾的。

在前面的几章里，我们详细地讨论了物质的起源问题。现已知道，恒星和行星之类的物体形成于原初气体，而宇宙物质本身则是在大爆炸中创生的。最近在粒子物理方面的发现，显现了宇宙引力场能够存空洞无物的空间里创造物质的一些机制。于是，未解之谜只剩下

了时空本身的起源。但是，有些迹象表明，空间和时间可以自发地产生而且不违反物理定律。之所以会有这种奇特的可能性，其原因涉及量子论。

我们说过，在亚核世界中量子因素准许没有原因的事件发生。例如，粒子可以在没有原因的情况下不知从什么地方冒出来。当量子论推广到引力时，就涉及时空本身的行为了。尽管现在还没有令人满意的量子引力理论，然而物理学家们对很多会被纳入这样一个理论的普遍现象有了相当的认识。例如，量子引力理论会把量子物质所特有的那种模糊的不可预测性赋予空间和时间。尤其应当注意的是，量子引力理论准许时空自发地、没有原因地创生和毁灭，就如同粒子自发地、没有原因地创生和毁灭一样。这一理论会包含着某种数学或然性，即一小块空间会从原先没有任何空间的地方出现。因而，没有原因的量子跃迁会使时空从无中突然产生出来。

通常，量子机制使时空突然出现的情况一般只发生在超微观尺度上，因为量子作用通常只适用于微观现象。实际上，自发产生的空间可能只有$10^{-33}$厘米大小。然而，这样小的有限空间不必有边缘，它可以是像第二章里所说的那样，是被关闭在超圆体之中的。很可能，这种微型宇宙会随着另一个相反的量子起伏而迅速消失。不过，也存在一种可能性，就是它非但没消失，反而像一个气球一样突然膨胀起来。

这种突然膨胀的行为，起源于其他的量子过程。这些过程与引力无关，却跟大自然其他的基本力有关。在第十三章里，我曾简短地描述了所谓的“暴涨宇宙”。其中我说，“大统一力”使初生的宇宙变得不稳定，从而使宇宙进入失控的指数膨胀的阶段。这样，量子微观世界就可能在不到1秒的时间里膨胀到宇宙的规模。积聚在这场大爆炸中的能量，就会随着暴涨阶段的突然结束而转化为物质和辐射，然后，宇宙就会像我们按常规所想的那样一路演化下来。

按这种不同寻常的宇宙创生理论来看，整个宇宙完全是从无中生出来的，其创生过程完全符合量子物理的定律。这样的宇宙在其成长过程中创造了一切物质、一切能量，从而建成了我们现在所看到的宇宙。这样，这种理论便说明了一切物理性的东西（包括时间和空间）的创生。这种量子时空模型并不规定一个不可知的奇点来给宇宙开头（见第二章），而是试图完全在物理定律的框架之内解释一切。这可是个让人敬畏的理论。我们习惯上都是认为“有失才有得”，还从没听说过能从无中得到什么东西。然而，量子物理却不当回事地从无中产生出一些东西来。量子引力表明，我们可以从无中获得一切。物理学家艾伦·古斯在讨论这一理论时说：“人们常说没有免费午餐这回事。然而，宇宙就是免费午餐。”①

这种从无中生出的宇宙还需要上帝吗？在第三章里，我们说过传统上的上帝存在的宇宙学论证如何要先设定一切事物都必有一个原因。量子物理学驳倒了这种设定。但是四大问题还剩下两个，即为什么宇宙会有今天的东西？为什么宇宙有今天的规律？科学能提供这两个问题的答案吗？

在第十一章里我们说过，所谓的超引力理论的目标，就是要提供一种数学描述方法，来描述大自然的一切力和物质的一切基本粒子。假如这一理论被证明能够成立，就可以把上面所说的那两个问题缩减成一个，组成世界的东西（质子、中子、介子、电子等）就可以在超引力理论的框架之内得到解释。现在的物理定律则与超引力理论不同。一般说来，有了电子或质子，我们就知道电子或质子会有什么行为，但我们却不知道为什么会有电子或质子存在，而不是有性质不同的其他粒子存在。假如超引力理论完全成功，那么，它就会不仅告诉我们为什么有各种粒子存在，而且能说明为什么各种粒子具有它们现有的质量、电荷以及其他的性质。

整个超引力理论要以一个宏伟的数学理论为基础，以使物理学的一切分支（还原论意义上的）都包括在一个超定律中。但是，我们最

后仍是有一个问题：为什么会有那个超定律？

这样，我们就碰上了有关存在的最终问题。物理学或许能够解释物质宇宙的内容、起源和组织，但却不能解释物理学本身的定律（或超定律）。按照传统看来，是上帝发明了大自然的所有定律，并创造了由这些定律所管辖的万物（时空、原子、人，等等）。“免费午餐”理论认为，只要有了各种定律，宇宙就会自己照管自己了，包括照管它自己的创生。

但是，定律又是从哪里来的呢？必须得先有规律，宇宙才能降生。必须得先有量子物理（从某种意义上说），量子跃迁才能产生出宇宙来。很多科学家认为，为什么物理定律是现在这样的，这个问题是无意义的，或者说，这个问题至少是不能给以科学的解答的。其他的科学家则持有“人存”观点，认为物理定律必须是现今这样的，否则就不会有观察者存在。但是，还有一种可能性。物理定律（或是最终的超定律）或许将显现出来，成为逻辑上唯一可能的物理原理。这就是我们要在最后一章里讨论的。

# 第十七章　物理学家心目中的大自然

大自然具有一种质朴性，因而非常优美。

理查德·费恩曼

假如你得到的只是质朴的美，

你得到的就是近乎上帝发明的最美的东西。

伊丽莎白·勃朗宁

在前面的那些章节里，我们探讨了新近的科学进展，尤其是人们所说的新物理学的进展对宗教的影响。尽管现代科学获得了辉煌的成功，但认定科学的进展已经解答了有关上帝存在的基本问题，解答了宇宙的目的以及人类在自然的和超自然的计划中的作用问题，这种想法则是愚蠢的。实际上，科学家们自己也有着广泛的宗教信仰。

经常有人说，科学和宗教可以和平共处，因为二者解决的是不同的问题。宗教教义的问题所涉及的是道德或三位一体的概念，等等。科学的问题则涉及判定何为引力的最佳数学描述，等等。两种问题有着本质的不同。然而不容否认的是，科学的确在宗教事物上有发言权。在时间的性质、物质和生命的起源、因果关系、决定论等问题上，宗教的概念框架可能随着科学的进展而更改。前几个世纪的一些重大宗教问题（如天堂和地狱的方位）已经随着现代宇宙学的出现，随着我们对空间和时间的性质的进一步了解而变得没有意义了。

很多人总喜欢用“正确与错误”的观念来打发科学与宗教的冲突。

人们很容易这样想：世上存在着终极的真理，这真理是一种客观的实在，科学和宗教都在探索它。按照这种显然合理的观念来看，“上帝存在吗？”“有超自然的奇迹吗？”“有天地创生这回事吗？”“生命起源于偶然吗？”“宇宙有目的吗？”这一些问题即使我们可能不明白，却都可以用“对”或“不对”来回答。

人们常常会碰上这样的观点：科学的理论都是近似真正的实在。随着我们的知识进步，理论与现实也会更加吻合。这种看法认为，大自然的“真正的”规律埋藏在观察和实验的数据之中，只有带着灵感进行坚持不懈的研究，才能把规律发掘出来；我们可以期望，到了将来的某一天，正确的规律将会被揭示出来，我们今日的教科书上的规律只是这些真正规律的可信然而却有缺陷的摹本。这在很多方面就是超引力理论所要达到的目标。超引力理论的支持者预期，他们将会发现一套方程式，将“真正的”规律完整地体现出来。

然而，并非所有的物理学家都认为谈论“真理”是有意义的。按照这样的观点来看，物理学所研究的根本不是什么真理，而是一些模型，即一些能帮助我们以一种系统的方式把一个观察与另一个观察联系起来的模型。尼耳斯·玻尔就表现出这种所谓的实证论的观点。他说，物理学家只是告诉人们我们能就宇宙了解些什么，而不是告诉我们宇宙如何存在。在第八章里我们说过，量子论已促使很多物理学家宣称，世上根本不存在“客观的”实在。那唯一的实在便是通过我们的观察而揭示出来的实在。假如承认这种观点是正确的，那就不可能说某个理论是“正确的”或“错误的”，而只能说某个理论是有用的或不大有用。某个理论有用，就是说它能高度精确地在一个单一的描述式中将范围广泛的多种现象联系起来。这样的观点是与宗教的观点截然相反的，因为宗教徒坚信有一个终极的真理。一般认为，一个宗教的命题不是对的就是错的，而不能像科学命题那样被看作是我们的经验的某种模式。

物理学家们总是乐于放弃某一个他们所喜爱的理论，以接受一个

更好的理论。这显示出宗教与科学在基本思想方法上的差异，正如罗伯特·默顿曾经写道的：“人类的大多数制度都要求人们绝对的信仰，而科学的制度则使怀疑成为一种美德。”当爱因斯坦发现了相对论时，人们意识到，牛顿的有关时空和力学的理论在描述以接近光速运动的物体行为方面是不完备的，于是，牛顿的理论便被取代了。牛顿的理论并不是错的，它只是适用范围有限。狭义相对论是一个更有用的理论（它使牛顿的理论变得只适用于低速运动），它能对高速系统进行更精确的描述。狭义相对论后来又为所谓的广义相对论所取代；而且，没有几个物理学家怀疑，广义相对论将来会被改进。说到某个“最终的”完美无缺的理论不可能再被改进，一些物理学家认为这种理论是无意义的。我们不能想象世上会有一幅完美无缺的图画，或一曲完美无缺的交响乐，同样，也不会有什么完美无缺不可改进的理论。

科学的方法能够随着新的科学发现而变更，这正代表了科学的伟大力量之一。通过使自己立足于实用而不是真理，科学便将自己与宗教显著区别开来。宗教是以教义和公认智慧为基础的，旨在代表不可变易的真理。尽管宗教教义的枝节问题可以随着时间改变，然而，要信教的人放弃宗教的基本教义去接受实在的一种更为“精确的”模型却是不可想象的。假如基督教会根据新的证据宣布，基督似乎并没有复活，那么，基督教就很难不变得面目全非。一些批评家说，教义僵硬意味着每一个新发现和每一个新观念都可能构成对宗教的威胁，而新的事实和新的观念则是科学的生命。因此，多少年来，科学的发现使得科学与宗教冲突不断。

尽管宗教是后顾天启真理，而科学则是前瞻新前景和新发现，但宗教和科学这两种人类活动都让参与活动的人感到敬畏，并在他们身上将谦卑和傲慢奇妙地混合起来。所有的伟大科学家都被他们要试图理解的自然界的精巧和优美所感动。每一个新的亚原子粒子，每一个未曾料到的天体都使他们感到惊喜。在构筑他们的理论的过程中，物

理学家们频频遵从神秘的优美观念的引导，相信宇宙具有内在的美。这种艺术趣味一次又一次地被证明是富有成果的指导性原理，常常能直接导致新的发现，甚至在乍看之下与观察的事实相悖的情况下也是如此。

保罗·狄拉克曾写道：

让方程式优美比让方程式符合实验更重要……因为差异可能是由于未能适当地考虑一些小问题造成的，而这些小问题将会随着理论的发展得到澄清。在我看来，假如一个人在进行研究工作时着眼于让他的方程式优美，假如他真有正常的洞察力，那么他就肯定会获得进步。①

玻姆用简洁的语言表达了相同的思想：“物理学是洞察力的一种形式，因而也是艺术的一种形式。”②

爱因斯坦在论述他不相信有一个人格的上帝的同时，又表达了他的赞美之情：“世界富于秩序与和谐，我们只能以谦卑的方式不完全地把握其逻辑的质朴性的美。”

在物理学家看来，美这一概念的关键是和谐、质朴、对称。我们再来看一段爱因斯坦的话：

一切科学工作都基于这种信仰，即存在应该具有一个完全和谐的结构。今天，我们比以往任何时候都更没有理由让我们自己人云亦云地放弃这个美妙的信仰。像引力场方程式这样复杂的方程式，只有通过发现一个具有逻辑质朴性的数学条件才能找到。③

爱因斯坦的这种心情近来得到了惠勒的回应：

物理定律的美，就是它们所具有的那种难以置信的质朴性……这一切背后的最终数学机件是什么呢？它肯定是最美的。④

今天，这种与艺术趣味相通的指导性原理正激励着人们寻求超引力。最近，两位倡导超引力理论的头面人物在评论超引力数学的进展时指出："所有的力都来自局部对称性的共同要求，从这里人们便能够瞥见一个能令人得到深刻满足的秩序。"⑤

当物理学家们谈论美和对称的时候，他们用以表达这些概念的语言是数学。数学对科学、尤其对物理学的重要性是怎样强调也不为过的。列奥纳多·达芬奇曾写道："人类的任何研究活动，假如不能够用数学证明，便不能称之为真正的科学。"今天的情况比起15世纪来或许更是如此。

大多数普通的人对数学总感到神经性的恐惧。这主要是由他们与物理学的隔阂造成的。这种隔阂便是一堵屏障，使他们不能充分地欣赏科学发现，也使他们在面对经历千辛万苦的研究才揭示出来的众多自然奇观时感觉不到快乐。正如罗杰·培根所说的那样："数学是进入各种科学的门户，是钥匙……没有数学知识，就不可能知晓这个世界中的一切。"⑥

许多物理学家深为大自然所具有的数学质朴性和大自然规律的优美所感动，以至他们认为，这种质朴性和优美所显示的正是存在的基本特点。詹姆斯·吉恩斯爵士曾经说，在他看来，"上帝是个数学家"。但是，为什么上帝会喜欢以数学形式将他的理念付诸实施呢？

数学是诗体的逻辑。若想表达某个规律，那么，以质朴的、不可摇撼的逻辑为基础的表达便是最足以服人、最令人满意的表达方式。用约翰·惠勒的话来说就是：

假如对大自然的描述最终把人们带到了逻辑那里，带到了数学中心处的梦幻般的城堡那里，这是没有什么好惊讶的。假如像人们所相信的那样，一切数学最终是逻辑的数学，一切物理学最终是数学，那么，一切物理学若最终不是逻辑的数学还会是什么呢？逻辑是数学的

唯一能够“思想自身”的分支。[7]

用逻辑来表达大自然的吸引力之一是，大自然在很大程度上（假如不是全部的大自然的话）有可能从逻辑推论而不是从经验证据中演绎出来。在第二次世界大战之前，亚瑟·艾丁顿和E·A·米勒尼都曾尝试建构关于宇宙的演绎性理论（但没怎么成功）。但他们的尝试引起了一个有趣的问题。这就是，宇宙之所以是今天这个样子，难道是不可避免的逻辑的必然吗？伟大的法国科学家让·达兰贝尔写道：“对一个从统一的观点把握宇宙的人来说，宇宙的整个创生过程看上去像是一个唯一的真实必然。”这种观念使上帝全能的问题变得有趣起来。在第十章里，我们曾经指出，一个全能的造物主能够造出他想造的任何宇宙。基督徒声称，之所以会有如今这个特定的宇宙可以用上帝的选择来解释。上帝出于一些不为我们所知的理由，从无尽的可能里选择了这个宇宙。但是，上帝虽然全能，也不能打破逻辑的规则。上帝不能使2等于3，或使一个正方形等于一个圆。谁要是急匆匆地假定上帝可以创造任何宇宙，都必然得改口说，上帝所创造的宇宙得符合逻辑。现在，假如说只存在一个合逻辑的宇宙，那么，上帝就没有什么选择可言了。爱因斯坦特别指出：“我所真正感兴趣的是，上帝能否用另一种方式创造这个世界，即逻辑质朴性的必然是否留出了任何的自由。”[8]

假如宇宙实际上只有一种可能的创生方式，我们为什么还需要一个造物主？这造物主除了“按电钮”使创生的过程开始之外，还起了什么作用呢？但是，这种按电钮的作用并不需要一个精神。那只是一种启动机制，而且我们在前一章里看到，在一个量子物理的世界里，连启动机制也不需要。这是否是说，认为宇宙的基本逻辑数学方程只有一个物理解便是否定了上帝的存在？确实没有否定。这种观点只是使造物主上帝这一观念成为多余的东西，却没有排除可能有一个普遍的精神，这精神属于这唯一的宇宙，是自然的而不是超自然的上帝。当然，就像我们自己的精神不处于空间一样，这里所说的“属于”不是“处

于空间某处”的意思。“属于”在这里的意思也不是说“是由原子构成的”，正如我们的精神（与大脑相对的）不是由原子构成的。大脑是人类精神的表达媒介。同样，整个的物质宇宙也是一个自然的上帝之精神的表达媒介。这样，上帝就是一个至高的整体概念，很可能要比人类精神高出很多描述层面。

假如承认这些观念，那么，了解物质宇宙的起源及命运就十分重要了。因为精神必须要有组织，所以，精神的存在就受到了热力学第二定律的威胁。随着宇宙慢慢地被它自己的熵闷死，上帝是不是也得死去？假如不是这样的话，那么就有另一种可能性，即宇宙发生引力崩溃，成为奇点，导致物质宇宙完全消失。这种可能性似乎更不美妙。只有宇宙是循环的或是稳态的，一个自然的上帝似乎才能有无限和永恒的机会。

到目前为止，我们所讨论的物理学家对自然的看法都是从还原论出发的。物理学家在寻找新的规律和模型的过程中深受大自然的美和质朴性的感动。这里所说的美和质朴性在很大程度上指的是那些构造了世界的基本结构：夸克和轻子之类的亚原子粒子，以及作用于这些粒子之间的各种基本力。但是，上帝的整体性再一次提醒我们，不管物理学家对世界的构造和质料了解得多么清楚，任何纯粹的还原论思路总是把握不了整体性特点的。

理查德·费恩曼曾这样表达了相同的意思：

按各种等级或层面谈论世界便是对世界进行讨论的一种方式。现在，我不想很精确，不想把世界划分为明确的层面，但是，我想通过描述一系列观点来表明我所说的观点的层次是什么意思。

例如，在一方面，我们有基本的物理定律。然后，我们创造了一些其他的概念术语，这些术语是不那么基本的，而且我们相信它们最终都可以由基本定律给以解释。比如说，“热”。热被认为就是轻轻跳

动。说一个东西热，就是说有一堆轻轻跳动的原子。但是，在一段时间里，当我们讨论热的时候，我们有时忘记了那些轻轻跳动的原子，正如我们谈论冰川的时候，我们并非总是想到冰川形成之前从天而降的六角形雪花和冰凌。另一个相同的例子是盐的晶体。从基本的角度来看，盐的晶体是许多质子、中子和电子。但是，我们却有“盐晶体”这个概念，这一概念携带了整个的基本相互作用的图式。压强的概念也是这样。

现在，假如我们往上去进入另一个层面，我们就有了物质的各种特性，如“折射率”，即光穿过某物时的弯曲程度；还有“表面张力”，即水将自己收缩起来的倾向。折射率和表面张力都是用数字描述的。我要提醒你们，我们得通过好几个定律的帮助才能发现张力就是原子的拉力，等等。但是，我们仍是漫不经心地说“表面张力”，在讨论表面张力时，并不总是把表面张力的内部作用挂在心上。

再往上的层次，我们就有了水的波浪，有了暴风雨这样的东西。“暴风雨”这个词代表的是一些现象。还有“太阳黑子”，“恒星”。恒星是由很多东西积累而成的。然而，总是要给这些东西追根溯源是不值得的。事实上我们也追溯不了，因为我们往上走的层次越多，我们和基本定律之间相隔的步骤也就越多，每一步骤都有点论据不足。我们还没把它们想通呢。

我们沿着这一复杂性的等级走上去，就碰上了肌肉痉挛或神经兴奋这一类的事物。这类事物是物质世界当中非常复杂的东西，因为涉及的是物质的非常复杂的组织。再往后，就有了“青蛙”这一类的东西。

我们再往上去，就碰上了“人”，“历史”或“政治权宜”之类的词语和概念。我们用这一系列的概念来理解处于较高层面的事物。

再继续往上走，我们就会碰上恶、美、希望……现在，我们一边有基本的定律，一边有美和希望。假如可以用一个宗教比喻的话，便

可以问：哪边更接近上帝？我想，正确的回答当然是说我们必须看到事物的完整的结构联系；我们必须看到，一切学科，不仅是一切学科，而是一切知识活动，都是要找出各个层次间的联系，要把美与历史联系起来，把历史与人的心理联系起来，把人的心理与大脑的活动，大脑活动与神经兴奋，神经兴奋与化学等这一切上下双向联系起来。今天，我们不可能（而且也没法假装相信我们能）仔细地划出一条线，把事物的这一边同另一边联起来，因为我们才刚刚开始发现有这样的相对等级。

我认为，两边距上帝都是一样远。⑨

在前面的那些章里我强调指出，科学家们现在越来越注意大自然的结构等级的重要性。像生命、组织、精神之类的整体性概念确实是有意义的，是不能用“不过”是原子、夸克、统一力或其他什么东西之类解释过去的。了解在一切自然现象的核心里都有基本的质朴性，这当然很重要。然而，不管它有多么重要，也不可能就是一切。复杂性和质朴性是同样重要的。

现代物理学的一个悬而未决的问题是，一个物理系统的整体性特点是否需要有另外的整体性规律，而这些整体性的规律是不能还原为基本力或粒子的基本规律的。到目前为止，我们没有任何真正的物理整体性规律存在的证据。例如，热力学处理的是气体之类的整体性系统，这些系统中有很多集体活动的分子。温度和压强之类的概念在单个分子的层面上是没有意义的。然而，所有的有关气体的定律都能够从低层面的分子运动规律得出来，只是这些运动规律以一种统计学的方式被用于众多的分子。真正的整体性规律应当是，在整体层面上显现出来一种新的力或具有组织作用的影响，这种力或影响并不是起源于单个的组成部分。这就是活力论在解释生命时的假设。

整体性物理定律的一个更为引人注目的例子是意念致动或心灵感应。对所谓的特异现象持肯定态度的人声称，人的精神实际上能够对

遥远的物质施加力。据认为，这样的力在还原性层面上是不为人们所知的。这些力不是核力，不是引力，也不是电磁力。这些意念力的最直接的表现是令人惊奇的远距离意念弯物，即被试者似乎在没有身体接触的情况下，全凭精神力量使金属物体变了形。笔者为了验证这一现象，设计了一个非常严格的试验，就是把金属棒封在玻璃器皿里，器皿中的空气被抽出，换上比例不为他人所知的多种稀有气体，以防止被试者搞鬼。最近，一些意念弯物的大师接受了测试，结果没有一个能使金属棒发生任何测得出来的变形。

前面曾经说过，物质世界的结构很可能部分地或全部是非常质朴的逻辑原理的作用结果，这些逻辑原理是以基本的数学形式表现出来的。然而，复杂性的问题为人们接受这一观点造成了困难。难道我们真能相信，比如说，生命和精神完全起源于逻辑规则，而不是起源于整体性的力？

最简单的逻辑规则的作用能产生出有趣而复杂的活动。这一点可以得到很漂亮的验证。剑桥的数学家约翰·康威发明了一个叫作“生命”的游戏。这游戏很简单，由一个人在一块被隔成很多小格（间隔）的板上玩。在一些小格里放上黑色的棋子，由这些棋子组成的图形按下面这套规则改变形状：

1.有2个或3个邻子的棋子在下一代（即下一步）仍然存活。

2.没有或只有1个邻近的棋子要“死”（于孤单），有4个或4个以上的邻子的棋子要死（于拥挤）。

3.若1个空格正好与3个不空的格为邻，那么，这个空格就要生出1个棋子来。

这些出生、存活、死亡的规则很简单。康威和他的同事发现，某些图形根据这些规则能演变出各种各样最令人惊讶的美丽图案。其

中，有两个图形系列最引人注目。第一个图形系列使人们看到，简单的形状可以演变为复杂的结构。我们可以看图25。图中的“种子”长成一朵花，花凋谢了，死了，留下了4个小“种子”。

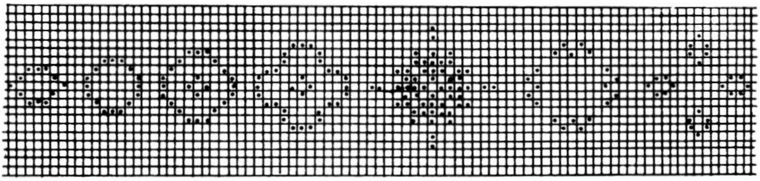


图 25　在约翰·康威的生命游戏里，有上面的一系列图形演变（有些中间步骤略去了）。这些图形恰巧像是一朵花的生命史。

更不同寻常的是，他们发现有些图形在演变过程中保留了某种连贯性，显示出一种很像是行为的活动。最简单的例子是“滑翔机”图形。这个图形连为一体在板上移动（见图26）。一些大的被称作“太空船”的图形在运动过程中会留下一串“火花”。然而，更大的“太空船”则需要一些较小的“护航船”来吃掉大船向前喷射的碎片，为大船开辟通道，否则碎片会把大船撞碎。

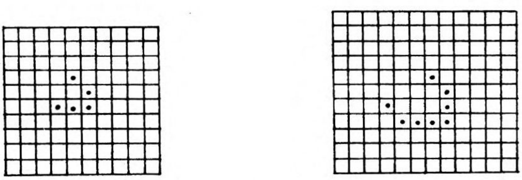


图 26　5 个小点简单地排列成“滑翔机”。这滑翔机有一个有趣的特征。它在板上作对角线运动，一直保持形状不变。由 8 个小点组成的图形叫“太空船”，但在运动时要喷射出“火花”。

借助于计算机，就可以用康威的游戏，来检验关于自复制的机器以及其他抽象的逻辑-数学难题的猜想。在检验中，先构造出一些图形，这些图形可以像生产线一样生出其他的图形。有一个图形叫“滑

翔机炮”，这种图形每30步就能产生出一个新的“滑翔机”。13个“滑翔机”碰撞之后产生的碎片就能够构成一个这样的“滑翔机炮”！仔细摆放“滑翔机炮”的位置，就能使相交的“滑翔机”的主干构成一个工厂，这工厂每300步就造出一个“太空船”。这一切“行为”都是“自动的”。只要有了选定的初始图形，这游戏自己就能产生出结构和行动，根本就不需要有人插手。而这一切结构和活动全来自几条简单的逻辑规则。

在我看来，物理学的重大贡献是通过还原论做出的。整体性的问题更适合于认知学科和科目，如系统论，形势运筹学，社会学，政治学。当然，这并不是说物理学跟整体论没有什么关系。物理学显然跟整体论有关系。热力学、量子论、自组织系统物理学都涉及整体概念。然而，我不认为物理学能处理目的或道德之类的问题。

有时候有人问我，物理学家通过研究大自然的基本过程得来的有关大自然的知识能否有助于理解上帝的宇宙计划的性质，或能否揭示出善与恶的斗争真相。我的回答是不能。夸克结合成为质子和中子，量子的被吸收或量子的放射，时空因物质产生的弯曲，将所有的基本粒子统一起来的抽象的对称，这一切的一切都谈不上什么善与恶。不错，在大自然里能发现很多竞争，例如，不同的力的平衡和相互作用之间就有竞争。一个恒星就是相互争斗的力的战场。引力要压垮恒星，因而就与要使恒星炸开的热压力和电磁辐射力发生争斗，而热压力和电磁辐射力又是核相互作用的能量释放产生的。宇宙中到处都在进行着这样的斗争。然而，假如对立的各种力不是大致旗鼓相当，所有的物理系统就会为这一种或那一种力所统治，宇宙的活动就会很快地停下来。宇宙中的这些斗争已持续进行了多少亿年，可见，宇宙是复杂的，有趣的。

这种斗争的僵局提供了极大的时间，使伟大的宇宙戏剧得以展开。对于那些看上去纯属偶然的“僵持”，弗里曼·戴森的话表达了人们的迷惑不解：

既然宇宙在一路滑向最后的死亡状态，在这死亡状态之中，能量的无序程度将会变为极大，然而，为什么宇宙竟能像查尔斯王一样死得这么费时？⑩

宇宙幸运地处于这种稳定状态，得以在很长的天文时间里不跌进完全的混沌之中。这种幸运的稳定是第十三章里讨论的各种“偶然”的一个方面。

有一种体积僵持，使宇宙得以避免在它自己的引力下突然崩塌。由于宇宙物质在空间中分布得极广，所以，宇宙发生爆聚时，宇宙间的物质由自由降落到形成大崩塌（假如大崩塌发生的话）所需的时间是上百亿年。还有一种旋转僵持，使星系和行星系稳定下来，不会收缩到它们自己的中心。离心力会来对抗引力向里的拉力。最后，还有核力的僵持，保持了恒星内部的核燃料消耗的速度非常缓慢。

这些僵持并不能世世代代永远持续不已。当僵持的局面维持不下去的时候，便经常爆发激烈的活动。宇宙充满了激烈的活动：恒星发生爆炸，躁动的星系和类星体喷射出巨大的能量，巨大的天体发生可怕的碰撞，天体被引力撕裂，物质在黑洞中被压缩得无影无踪。这一切，都是可怖的激烈活动。然而，物理学家在这些激烈的活动里却看不出有什么恶。在这里能量的肆行无忌造成的宇宙大乱中，大自然可能播下来安宁的种子。构成我们这颗恬静的星系的重元素，就是很久以前在超新星爆炸的大火中创生的。在物理学家看来，激烈活动的现象只是大自然规律的一种特殊的表达方式而已，从道德上讲是没有什么善恶可言的。善恶只能适用于精神，不能适用于物质。

在前面的那些章里，我们涉猎了整个现代物理学。我们为了寻找上帝，探讨了关于空间与时间的新观念，探讨了人们如何重新认识有序和无序，精神与物质。我们所说过的很多东西，无疑会证实某些人的想法，使他们更加坚信科学与宗教是死对头，科学会继续危及大部分宗教教义的基础。很多关于上帝、人以及宇宙本质的宗教观念已被

新物理学所破除，否认这个事实是愚蠢的。但是，我们在寻找上帝的过程中，也发现了很多确实的东西。例如，我们知道了宇宙中存在着精神，精神是一种抽象的、整体的组织模式，甚至可以离体存在。于是，还原论者认为我们不过是一堆堆活动的原子的看法就受到了反驳。

然而，我写这本书的意图并不是要为长期悬而未决的宗教问题提供简易的答案。我的目的是扩展对传统的宗教问题的讨论框架。新物理学已经推翻了很多关于空间、时间和物质的常识性概念，因而严肃的宗教思想者不能忽视新物理学。

在本书的开头，我曾声言，与宗教相比，科学能为寻找上帝提供一条更为切实的途径。我深信，只有从各个方面全方位地了解世界，从还原论和整体论的角度，从数学和诗的角度，通过各种力、场、粒子，通过善与恶，全方位地了解世界，我们才能最终了解我们自己，了解我们的家——宇宙背后的意义。

# 精选参考书目

第一章　科学与宗教在变化着的世界中

很多科学家写了论及他们自己的宗教信仰的著述，很多神学家也详细探讨了科学对宗教的影响。属于前一类的书有：

W. Russell Hindmarsh,Science and Faith（Epworth 1968）（《科学与信仰》）；

Donald Mackay,The Clockwork Image（Inter-Varsity Press 1974）（《钟表形象》）；

Jacques Monod,Chanceand Necessity（Random House 1971）（《偶然与必然》）；

C. A.Coulson,Science and Christian Belief（Oxford University Press）（《科学与基督教信仰》）；

Jacob Bronowski,Science and Human Values（Harper and Row 1965，revised edition）（《科学与人类的价值观》）；

还有Physics,Logic,History（eds.Allen D.Breck and Wolfgang Your-grau}Plenum 1970）（《物理学、逻辑、历史》）。

属于后一类的书有：

Stanley Jaki；Cosmosand Creator（Scottish Academic Press

1981）（《宇宙与造物主》）；

Hugh Montefiore；Science and the Christian Experiment（Oxford Uni-versity press 1971）（《科学与基督教的试验》）；

Thomas Torrance,Theological Science（Oxford University Press 1981）（《神学的科学》）；以及Divine and Contingent Order（Oxford University press 1981）（《具有神性的偶然秩序）：

R. Hooykaas,Religion and the Rise of Modern Science（Erdmans 1972）（《宗教与现代科学的兴起》）。

The Sciences and Theology in the Twentieth Century（ed. A.R.Pea-cocke；Oriel 1981）（《二十世纪的科学与神学》）是一本有趣的论文集。

第二章　创世

宇宙的创生近来一直是大量科普书籍的主题。其中最广为人知的是：

Steven Weinberg,The First Three Minutes（Andre Deutsch 1977；Fon-tana 1978）（《最初三分钟》）。

亦参看：

P. W.Atkins,The Creation（Freeman 1981）（《宇宙的创生》）；

Tohn Gribbin,Genesis（Dent/Delacorte 1981）（《创世》）

Joseph Silk,The Big Bang（Freeman 1980）（《大爆炸》）。

我在我写的TheRunaway Universe（Dent 1978；Harper&Row 1978）（《失控的宇宙》）一书中，详细地探讨了宇宙的创生及终结。该书强调了热力学第二定律的意义。与该书主题相同的另一本书是：

Robert Jastrow,Until the Sun Dies（Norton 1977）（《直到太阳死灭》）。

若想得到关于现代宇宙学的直截了当的综合述评，请看：

The State of the Universe（ed. G.T.Bath；Oxford：Clarendon Press 1980）（《宇宙的状态》）；

E. R.Harrison,Cosmology（Cambridge University Press 1981）（《宇宙学》）；或D. W.Sciama,Modern Cosmology（Cambridge University Press；Second edition 1982）（《现代宇宙学》）。

关于爱因斯坦的相对论，有很多通俗的解说。

Nigel Calder,Einstein's Universe（B. R.C.Publication 1979）（《爱因斯坦的宇宙》）是一部虽然浮泛却还生动的总评。更具学术性却依然通俗的著作有：

Peter Bergmann,The Riddle of Gravitation（Charles Scribner 1968）（《万有引力之谜》）；

Robert Wald,Space,Time and Gravity（University of Chicago Press 1977）（《空间、时间和引力》）；以及本作者的：

Space and Time in Modern Universe（Cambridge University

Press）（《现代宇宙的时与空》）。

若是有兴趣真正了解一些相对论，那么，纯粹的初学者可以看Sam Lil-ley,Discovering Relativity for Yourself（Cambridge University Press 1981）（《自己发现相对论》）；以及Robert Ceroch,General Relativity from A to B（University of Chicago Press 1978）（《广义相对论初步》）；

有点数学基础的则可以看：

Michael Berry,Principles of Cosmology and Gravitation（Cambridge U-niversity Press 1976）（《宇宙学的原理和万有引力》）。

无限的问题是普及数学的好题目，可以看Rozsa Peter,Playing with Infinity（Bell 1961）（《玩赏无限》）；

Leo Zippin,Uses of Infinity（Random House 1962）（《无限之用途》）；更晚近一些的书则有Ruddy Rucker,Infinityand the Mind：The Science and Philosophy of the Infinite（Harvestee.1982）（《无限与精神：关于无限的科学与哲学》）。亦参看“数学解无限”（by Keith Devlin in New Scientist，95，162）。

第三章　上帝创造了宇宙？

关于粒子和力的现代理论（包括有关反物质和粒子创生的讨论）的简介，请看我的：The Forces of Nature（Cambridge University Press 1979）（《自然的诸种力》）；以及Gerald Feinberg,What Is the World Made of？（Doubleday 1977）（《世界是由什么组成的》）

关于统一力的理论及质子衰变，请看Steven Weinberg，“质子的衰变”，（Scientific American,June 1981）。

关于上帝存在的宇宙论论证的深入的评论，请看：William Rowe,The Cosmological Argument（Princeton University Press 1975）（《宇宙论的论证》），以及William Lane Craig,The Cosmological Argument from Plato to Leibniz（Macmillan 1980）（《从柏拉图到莱布尼茨的宇宙论论证》）。参看：D.R.Burrill,The Cosmological Argument：A Spectrum of Opinion（Doubleday-Anchor 1967）（《宇宙论的论证：一系列的观点》）。

罗素与柯普莱斯顿的论战发表在：

A Modern Introduction to Philosophy（eds. P.Edwards and A.Pap；Free Press 1965）（《哲学的现代简介》）。

从宇宙学角度对宇宙的开初进行的详细历史探讨见于：

J. D.North,The Measure of the Universe（Oxford.Clarendon Press 1965）（《宇宙的测度》）。

第四章　为什么会有个宇宙？

关于时间之外的上帝这一概念的探讨，请看Richard Swinburne,The Coherence of Theism（Oxford：Clarendon Press 1977）（《有神论的一致性》）。亦参看：

Nelson Pike,God and Timelessness（Routledge&Kegan Paul 1970）（《上帝与无时间性》）；以及Brian Davies,An Introduction to the Philosophy of Religion（Oxford University Press 1982）（《宗教哲学入门》）的第八章。

什么是必然的存在？请看前面所说的S winburne的著作以及John Hosper,An Introduction to Philosophical Analysis（Routledge&Kegan Paul,revised edition 1981）（《哲学分析入门》）第七章。他是从怀疑论出发对这一问题进行探讨的。

David Layzer在《科学美国人》（Scientific American）1975年12月一期上的文章“时间之矢”也探讨了秩序如何能从原初的无序中产生出来的问题。Steven Frautschi在“膨胀的宇宙中的熵”（即将发表）一文中论述了具体的数学分析。

我的The Physics of Time Asymmetry（Surrey University Press/U-niversity of California Press 1974，1977）（时间不对称的物理学）一书探讨了引力在产生宇宙秩序过程中的作用问题，以及现在尚未解决的引力熵的问题。探讨同样问题的还有Roger Penrose的“奇点与时间不对称”一文，见General Relativity：an Einstein Centenary Survey（eds.S.W.Hawking and W.Israel；Cambridge University Press 1979）（广义相对论：爱因斯坦百年综述）一书，亦见P.C.W.Davi es的“热力引力是通向量子引力的途径吗？”一文，收入Quantum Gravity2：A Second Oxford Symp osium（eds.C.J.Isham,R.P enrose and D.W.Scia-ma；Oxford：Clarendon Press 1981）（《量子引力2：第二次牛津讨论会》）

第五章　什么是生命？整体论对还原论

关于生命的性质及起源，有很多引起争议的观点。下列的书单是随便造的：

A. G.Cairns-Smith,TheLife Puzzle（Oliver&Boyd 1977）（《生命之谜》）；

Francis Crick,Life Itself：Its Origin and Nature（Macdonald/

Smith&Schuster 1982）（《生命：起源与性质》）；

Richard Dawkins,The Selfish Gene（Oxford University Press 1977）（《自私的基因》）；

Gerald Feinberg and Robert Shapiro,Life Beyond Earth（William Mor-row 1980）（《天外的生命》）；Fred Hoyle and N.C.Wickramasinghe；Lifecloud（Dent 1978）（《生命云》）；Jacques Monod；Chance and Ne-cessity（Random House 1971）（《偶然与必然》）；以及：L.E.Orgel,The Origin of Life：Molecules and Natural Selection（Wiley 1973）（《生命的起源：分子与自然选择》）。

关于机械论-活力论问题的探讨，见：

Hans Driesch,The History and Theory of Vitalism（Macmillan 1914）（《活力论的历史和理论》）；Rainer and Schubert-Soldern,Mechanism and Vitalism（Notre Dame University Press 1962）（《机械论与活力论》）。

用“蚂蚁赋格曲”和其他清楚明白的方式对整体论-还原论问题进行的解说，见于：

Douglas Hofstadter,G？del,Escher,Bach（Basic Books 1979）（《哥德尔，埃舍尔，巴赫》）。

对物理学中的还原论进行的猛烈攻击见于：

Fritjov Capra,The Tao of Physics（Wildwood House 1975）（《物理学的道》）以及The Turning Point（Wildwood House 1982）（《转折点》）。亦参看：

Paul Reps,Zen Flesh,ZenBones（Penguin 1971）（《禅肉，禅骨》）。

Gary Zukav,The Dancing Wu Li Masters：An Overview of the New Physics（Rider 1979）（《跳舞的物理大师们：新物理学总览》）一书也讨论了在现代整体论物理学中某些东方神秘主义的意味。

对自组织的化学和生物系统的深入研究见于：

G. Nicolis and I.Prigogine,Self-Organization in Non-equilibrium Sys-tems（Wiley 1977）（《非平衡系统中的自组织》）。从一更为广泛的背景出发对同一问题进行研讨的有：I.Prigogine,From Being to Becoming（Freeman 1980）（《从存在到变化》）；以及：Hermann Hake n,Syner-getics（Springer 1977）（《协同学》）。

天外生命，无论是有智能的还是没有智能的，都是科幻和科普作品的热门话题。然而，现在科幻与科普之间的区别不总是那么清楚。属于科普类的书有：

Ronald Bracewell,The Galactic Club：Intelligent Life in Outer Space（Freeman 1974）（《星系俱乐部：宇宙空间里的有智能的生命》）；前面已说过的Feinberg和Shapiro的著作；前面已说过的Hoyle和W ickramas-inghe的著作；Interstellar Communication：Scientific Perspectives（eds.C.Ponnamperuma and A.G.W.Cameron Houghton-Mifflin 1974）（《星际通讯：科学的展望》）；

Carl Sagan,The Cosmic Connection（Doubleday 1973）（《宇宙间的联系》）；

I. s.Shkovskii and C.Sagan,Intelligent Life in the Universe（Hold-en-Day 1966）（《宇宙中的有智能的生命》），以

及

Walter Sullivan,We Are Not Alone（Mc Graw-Hill 1966）（《我们并不孤单》）。

天外生命在宗教上会有什么含义？对此问题进行的思考见C.S.Lewis的“宗教与火箭学”一文，该文收在The World's Last Night and Other Essays（Harcourt Brace Jovanovich Inc.1952）（《世界的最后一夜及其他》）。亦参看同一作者的Perelandra（Macmillan 1944）。

第六章　精神与灵魂

哲学和心理学的中心问题都是精神。若想对精神的许多方面有一个全面的概括了解，请看：

Richard Gregory,Mind in Science（Weidenfeld&Nicolson 1981）（《科学中的精神》）；

对精神问题进行了一番更为随和而具有思辨性的综述的是The Mind's I（eds.D.R.Hofstadter and D.C.Dennett；Harvester/Basic Books 1981）（《精神的我》）。上面的两本书都附有详尽的参考书目。亦参看：

D. C.Dennett,Brainstorms（Bradford Books 1978）（《脑猝变》）；以及Oxford Companion to theMind（ed.R.L.Gregory；Oxford Universi-ty Press 1983）（《牛津精神侣伴》）。

有关人工智能的讨论，见：

Philip Jackson,Introduction to Artificial Intelligence（Petrocelli Char-ter 1975）（《人工智能导论》），此书附有详备的参考书目；

Pamela Mc Corduck；Machines Who Think（Freeman 1979）（会思想的机器）

Philosophical Perspectives in Artificial Intelligence（ed.M.Ringle；Humanities Press 1979）（《人工智能的哲学展望》）；以及

Patrick Winston,Artificial Intelligence（Addison-Wesley 1977）（《人T智能》）。亦参看：

Douglas Hofstadter,C?del,Escher,Bach（Basic Books 1979）（《哥德尔，埃舍尔，巴赫》）

就意识问题，若想得到一个通俗的简介，可以看：

Eric Harth,Windows on the Mind：Reflections on the Physical Basis of Consciousness（Harvester 1982）（《精神之窗：关于意识的物理基础的思考》）

到了现代，仍为二元论辩护的有：

Karl Popper and John Eeeles,The Self and its Brain（Springer 1977）（《自我和它的大脑》）；以及H.D.Lewis,Philosophy of Religion（The English Universities Press 1965，1975）（宗教哲学）。

第七章　自我

自我和个人同一性之谜长时间以来一直为哲学家所关注。最近的有关著作有：

Bernard Williams,Problems of the Self（Cambridge University

Press 1973）（《自我的问题》）；

Pe rsonal Identi ty（ed. John Perry；Unive rsity of Cal ifornia Press 1975）（《个人同一性》）；以及

Sydney Shoemaker,Self-Knowledge and Self-Identity（Comell Univer-sity Press 1963）（《自知与自我同一性》）。

旨在让不熟悉形式逻辑的人也能理解哥德尔的有名的定理的一本书是：

Ernest Nagel and James R. Newman,G？del's Proof（New York Uni-versity Press 1958）（《哥德尔的证明》）。

讨论精神与大脑时也强调层面区分的重要性的一本书是：

Hierarchy Theory；The Challenge of Complex Systems（ed.H.H.Pattee；George Braziller 1973）（《等级理论：复杂系统的挑战》）。

有两本书论述了大脑移植以及因而造成的身份混淆，其中一本是前面说过的Perry的著作，另一本是The Identites of Persons（ed.A.O.Rorty；University of California Press 1976）（《人的身份》）。

第八章　量子因素

有关量子论的普及读物很罕见，我的Other Worlds（Dent 1980）（《其他的世界》）是我尽了最大的努力写成的。强调量子的整体性特点的有David Bohm,Wholeness and the Implicate Order（《整体性与隐含的秩序》）。若想得到有关问题的通俗简介，可以看：

Banesh Hoffmann,The Strange Story of the Quantum（Dover,second edition 1959）（《量子的奇异故事》）。对那些具有专业知识的人来说，若想就有关认识论的问题得到一个较好的综述，可以看Bernard d'Espag-nat,Conceptual Foundations of Quantum Mechanics（Benjamin 1971）（《量子力学的概念基础》）。可以先看看该作者的“量子论与实在”，见于《科学美国人》1979年第11期。

量子论创始人自己所撰写的有关经典有：

Niels Bohr,Atomic Theory and the Description of Nature（Cambridge University Press 1934）（《原子理论和对大自然的描述》）；

Werner Heisenberg,The Physical Principles of Quantum Theory（Uni-versity of Chicago Press 1930）（《量子论的物理原理》）；

Physics and Philosophy（Harper&Row 1958）（《物理与哲学》）；

The Physicist's Conception of Nature（Hutchinson 1958）（《物理学家心中的大自然》）；Max Born,Natural Philosophy of Cause and Chance（Oxford University Press 1949）（《原因和偶然的自然哲学》）。

Quantum Theory and Beyond（ed.T.Basin；Cambridge University Press 1971）（《量子论及未知》）一书里有很多有用的概念。

对Aspect跟他人合作的实验的描述，见：Physical Review Letters 49，1804（1982）.

John Wheeler在“黑洞之外”一文中描述了他的“延迟选择”实验。该文收在：

Some Strangeness in the Proportion：ACentennial Symposium to cele-brate the Achievements of Albert Einstein（ed. H.Woolf；Addison-Wesley 1980）（《比率中的某种异常：庆祝爱因斯坦的伟大成就百年纪念学术讨论会》）

对量子测量问题的分析，见：

J.yon Neumann,Mathematical Foundations of Quantum Mechanics（Princeston University Press 1955）（《量子力学的数学基础》）。

Everett对量子论的解释，其详尽的说明见：B.S.Dewitt and N.Gra-ham,The Many-Worlds Interpretation of Quantum Mechanics（Prince-ton University Press 1973）（《量子力学的多世界解释》）。若想得到一个较为易懂的简介，请看Dewitt的“量子力学和实在”，收入1970年9月1期的Physics Today。

头一次证明贝尔的不等式的是Physics1，195（1964）上发表的论文。

第九章 时间

多年来，时间一直是物理学、哲学和神学中一个吸引人的问题。有关时间的文献汗牛充栋。若想初步了解关于时间的物理学，可以看我的Space and Time in Modern Universe（Cambridge University Press 1977）（《现代宇宙中的时与空》）；

G.J.Whitrow,The Natural Philosophy of Time（Nelson 1961）（《时间的自然哲学》）；或者阅读任何一本关于相对论的入门书

（《参看第二章的参考书目》）。

侧重于时间之“矢”的物理学和哲学的书有：

P.C.W.Davies,The Physics of Time Asymmetry（Surrey University Press/University of California Press 1974，1977）（《时间不对称的物理学》），该书附有很多参考资料；

The Nature of Time（ed.T.Gold；Cornell University Press 1967）（《时间的本质》）；

Hans Reichenbach,The Direction of Time（University of California Press 1956，1971）（《时间的方向》）；The Philosophy of Space and Time（Dover 1958）（《时空的哲学》）。

若想了解一下哲学和神学大致如何看待时间，可以看下列的书：

The Concepts of Space and Time（ed.M.Capek；Reidel 1976）（《时间和空间的概念》）；

M. Capek,Philosophy of Space and Time（Dover 1958）（《时间和空间的哲学》）；

J.T.Fraser,TheGenesis and Evolution of Time（University of Mas-sachusetts Press 1982）（《时间的起源和演化》）；

The Study of Time（eds.J.T.Fraseret.a1.；Springer）（《时间研究》），卷2（1975），卷3（1978）；

Philosophy of Time（eds. E.Freeman and W.Sellars；Open Court 1971）（《时间哲学》）；

Time,Reduction and Reality（ed.R.Healey；Cambridge University ；Press 1981）（《时间，衰减与实在》）；

Nelson Pike,God and Timelessness（Routledge&Kegan Paul 1970）（《上帝与无时间性》）；

L.Sklar,Space,Time and Spacetime（University of California Press 1974）（《空间、时间与时空关系》）；

Space and Time（ed. J.J.C.Smart；Macmillan 1964）（《空间与时间》）；

Richard Swinburne,Space and Time（Macmillan 1968）（《空间与时间》）。

论述“现在”所包含的问题和悖论的有：

A.Grünbaum,Philosophical Problems of Space and Time（Alfred A-Knopf 1964）（《空间与时间的哲学问题》）；

J.McT.E.Mac Taggart发表在Mind 18，457（1908）上的“时间的非实在”一文，表明了他试图完全拒斥时间概念。

若想轻松愉快地探讨这类问题，John Gribbins的Timewarps（Dent 1979）（《时间弯曲》）是一本好书。

若想进一步阅读有关黑洞问题的书，请看第十三章的参考书目。

第十章　自由意志与决定论

前面已提到的Hospers的著作对自由、宿命论、决定论和非决定论作了明白易懂的讨论。

已有若干书专门讨论这类问题，如：

Free-will and Determinism（ed.B.Berofsky；Harper&Row 1966）（《自由意志与决定论》）；

Determinism and Freedom in the Age of Modern Science（ed.S.Hook；New York University Press 1957）（《现代科学时代中的决定论与自由》），以及Freedom and Determinism（ed. K.Lehrer；Random House 1965）（《自由与决定论》）。

强调上帝行动自由的书有Swinburne的The Coherence of Theism（Oxford：Clarendon Press 1977）（《有神论的一致性》）。

第十一章　物质的基本结构

对现代粒子物理学的入门性综述包括：

Nigel Calder. The Key to the Universe（Viking Press 1977）（《宇宙的答案》）；

J.S.Trefil,From Atoms to Quarks：An Introduction to the Strange World of Particle Physics（Charles Scribner 1980）（《从原子到夸克：粒子物理奇异世界入门》）；

J.C.Polkinghorne,The Particle Play（Freeman 1980）（《粒子波动》）；以及第三章的参考书目里列出的Feinberg和我的著作。

粒子物理学以及基本力的理论现在正在迅速发展，所以若想了解其概况，最好的办法是阅读《科学美国人》上定期刊载的评论文章。我想推荐的文章如下：

S.L.Glas how的“具有色、味的夸克”（1975年10月）；Y.Nambu的“夸克的密封”（197 6年11月）；S.Weinberg的“基本粒子相互作用的统一理论”（1974年7月）。若要进一步，连所谓大统一理论也包括进来的话，可阅读Howard Georgi的“基本粒子和各种力的统一理论”（1981年4月）。

D. Z.Freedman和P.yon Niewenhuizen的“超引力与物理诸定律的统一”评论了最终的理论——超引力理论（1978年2月）。这些文章有一部分被收入Partic les and Fields（ed.W.J.Kaufmann III；Free-man 1980）（《粒子与场》）。

Ed ward Kolb和Michael Turner的“早期宇宙”（发表于Nature 294，521〔1981〕）总结了人们近来将上面所说的思路应用于研究极早期宇宙的一些想法。

Capra的The Tao of Physics（《物理学的道》）强调了粒子的整体性方面，即粒子是由其他的粒子“组成的”。

第十二章 偶然还是设计？

我的The Accidental Universe（Cambridge University Press 1982）（《偶然的宇宙》）一书论述了大自然中一系列明显的“巧合”或“偶然”。与我的书主题相同的还有John Barrow和Frank Tipler的The Anthropic Principle（Oxford University Press）（《人存原理》），他们的书里还有有趣而深入的历史分析。

我在The Great Ideas Today（Encyclopedia Britannica 1979）（当今的伟大思想）里撰写了“宇宙中的有序和无序”一文，总结了宇宙中的有序无序的问题。亦请参看我的The Runaway Universe（Dent 1978；Harper&Row 1978）（《失控的宇宙》）。

对Poincaré循环的讨论，见Reichenbach的The Direction of

Time，亦见我的：The Physics of Time Asymmetry.

对循环宇宙的模型进行了分析的有；

R.Tolman,Relativity,Thermodynamics and Cosmology（Oxford.Clarendon Press 1934）（《相对论，热力学和宇宙学》）；循环宇宙的模型也成为John Wheeler讨论宇宙“重新处理”的基础。他的讨论见：

C.W.Misner,K.S.Thorne and J.A.Wheeler,Gravitation（Freeman 1973）（《万有引力》）第44章。

最近的文献中出现了若干讨论“人存原理”的，如B.J.Carr和M.J.Re es的“人存原理与物质世界的结构”（Nature 278，6 05，〔1979〕），以及I.L.Rozental的“物理定律与基本常数的数值”（Sovi-et Physics〔Uspekhi〕23，296〔1980〕）。这两篇文章很好地介绍了有关的物理学问题。John Leslie在American Philosophical Quarterly（美国哲学季刊）19，141（19 82）上发表的“人存原理，世界全体与设计”一文，探讨了人存原理的一些哲学问题。亦参看John Bowker的“上帝创造了宇宙？”收入The Sciences and Theology in the Twentieth Century（ed.A.R.Peacocke；Oriel 1981）（《二十世纪的科学与神学》）。当然，也应参看前面提到过的Barrow和Tipler的著作。若想得到一个入门性的概要，可以看George Gale的“人存原理”，发表在（《科学美国人》）1981年12月1期上。

第十三章　黑洞与宇宙混沌

黑洞是科普的热门题材，但应小心为是，因为这一题材常常处理不当。对此问题进行精心而又吸引人的探讨的有：

Larry Shipman,Black Holes,Quasars and the

Universe（Houghton-Mif-flin 1976）（《黑洞，类星体和宇宙》）；

Iain Nicolson,Gravity,Black Holes and the Universe（David&Charles 1981）（《引力，黑洞和宇宙》）；

Willam Kaufmann,Black Holes and Warped Spacetime（Freeman 1979）（《黑洞与弯曲的时空》）；以及那些最通俗的论述相对论的书（见第二章的参考书目）。我在The Edge of Infinity（Dent 1981）（《无限的边缘》）一书中深入地探讨了黑洞问题。

对引力熵这一概念的讨论，以及引力场对宇宙中的秩序和无序的重要作用，见我的《时间不对称的物理学》一书。近年来，Roger Penrose大大发展和澄清了这一科目，见他的“奇点与时间不对称”一文，收入General Relativity：an Einstein Centenary Survey（eds.s.W.Hawk-ing and W.Israel；Cambridge University Press 1979）（《广义相对论：爱因斯坦百年综述》）。这本书里也收有R.H.Dicke和P.J.E.Peebles的文章“大爆炸宇宙学——谜与妙方”，讨论的是视界问题以及相关的其他问题。

J.D.Barrow和R.A.Matzner在“宇宙的均衡性与各向同性”一文中，探讨了宇宙原初剧烈非均匀运动的耗散所造成的过量热量的难题，该文见于Monthly Notices of the Royal Astronomical Society 181，719（1 977）。亦参看Ba rrow的“静止宇宙学”一文，发表于Nature 272，211（1978）。John Barrow向我指出，即使质子真会衰变，原初宇宙的某些非均匀的东西仍会在后期宇宙中留下印记，即空间的非均匀。他在与M.s.Tur ner合作的论文中讨论了这一问题，见Nature 291，469（1981）。

Freeman Dyson的文章“宇宙中的能量”，讨论了核力若有一点点

变化就会造成什么样的灾难性后果的问题，见《科学美国人》1971年9月。

上帝存在的目的论论证，长期以来一直是有人赞同也有人不赞同。阿奎那修改了亚里士多德的第一推动说，提出了最终原因的概念，认为最终原因便是世界上存在神意的证明。然而，在阿奎那之后很久，直到出现了诸如Paley和F.R.Tennant的Philosophic al Theology（Cam-bridge University Press 1969；初版1928）（哲学神学）这样的著作，目的论论证才完善起来。康德的《纯粹理性批判》，以及休谟的《关于自然宗教的对话》都批驳了目的论论证。

现代对这一问题进行讨论的有：

Thomas McPherson,The Argument from Design（Macmillan 1972）（从设计出发进行的论证）；Rich ard Swinbur ne在Philosophy 43，200（1968）上写的一篇题目相同的文章。Bowker在前面提到的著作里叙述了近来有人试图把目的论改造成为“目的论天文学”。

第十四章　奇迹

休谟在其Enquiry Concerning Human Understanding（《人类理智研究》）一书的第十章，对奇迹的问题提出了或许是最著名的评判。

在当代对奇迹进行评判的有：

R.H.Fuller,Interpreting the Miracles（S.C.M.1966）（《解释奇迹》）；

R.Swinburne,The Concept of Miracle（Macmillan 1970）（《奇迹的概念》）；

Brian Davies,An Introduction to the Philosophy of Religion（Oxford University Press 1982）（《宗教哲学导论》）。

J.G.Taylor是位先前相信奇迹的科学家，后来他从怀疑论的观点写了一本综述所谓特异现象的书Science and the Supernatural（M.T.Smith 1980）（《科学与超自然现象》）。

第十五章　宇宙的终结

我在《失控的宇宙》一书中详细地探讨了这一问题。近来有若干文章、论文专门讨论这一问题，如J.N.Islam的“宇宙的最终命运”，见Sky and Telescope 57，13（1979年1月）；D.N.Page和M.R.Mc-kee的“永恒事关重大”，见Nature 291，44（1981），J.D.Barrow和F.J.Tipler的“永恒是不稳定的”，见Nature 276，453（1978）.Barrow认为，尽管宇宙物质会堕入热死，然而宇宙引力场中的扭曲可以被宇宙膨胀速率中细微的不规则所持续维持。引力场的扭曲可被用作自由能量的不尽来源。亦参见M.J.Rees的“宇宙的崩溃：末世学研究”，见Observatory 89，183（1969）。

不断膨胀的宇宙会越来越接近热死，在这样的宇宙中智能生命能否生存下去？Freeman Dyson在“没有终结的时间：开放宇宙的物理学和生物学”一文中详细探讨了这令人困惑的问题，见Review of Modern Physics 51，447（1979）。

W.Ehrenberg的“麦克斯韦妖”一文，介绍了麦克斯韦妖是怎么回事，见《科学美国人》1967年11月1期。L.Szilard的论文“论智能生物使一热力学系统的熵减少”发表于Zeitschrift fiir Physik 53，840（1929）。第一次全面论述了“魔鬼研究”中的信息和熵的联系问题的，是Leon Brilluin的论文“麦克斯韦妖成不了事”见Journal of Applied Physics 22，334（1951）。

## 第十六章　宇宙是“免费午餐”吗？

宇宙可能是某种量子起伏吗？若干物理学家探讨过这个问题。所提出的那些不大惊人的理论认为，构成宇宙的量子起伏来源于空的、平坦的时空，见E.P.Tryson的“宇宙是真空起伏吗？”发表于Nature 246，396（1973）；R.Brout,F.Englert和E　Gunzig的“作为一种量子现象的宇宙创生”，发表于Annals of Physics 115，78（1978）；D.Atkatz和H.Pagels的“作为量子穿越隧道事件的宇宙起源”，发表于PhysicalReview D 25，2065（1982）。

那些较为惊人的理论则认为，构成宇宙的量子起伏完全来自无。苏联物理学家Ya.B.Zeldovich在Soviet Astronomy Letters 7，579（1981）发表的一篇论文就这么认为。亦见L.P.Grishchuk和Ya.B.

Zeldovich的“完全的宇宙学理论”，收入The Quantum Theory of Space and Time（eds.M.J.Duff and C.J.Isham；Cambridge University Press 1982）（《时空量子论》）；以及A.Vilenkin的“宇宙创生于无”，发表于Physics Letters B 117，25（1982）。

对暴涨宇宙说的评论，见J.D.Barrow和M.S.Turner的“暴涨宇宙——诞生，死亡及变形”，发表在Nature 298，801（1982）。

## 第十七章　物理学家心目中的大自然

要了解物理学家如何看大自然，最好的方法是与他们交谈。可以当作指南的一本书是A Question of Physics：Conversations in Physics and Biology（eds.P.Buckley and F.Peat；Routledge&Kegan Paul 1979）（《物理学的一个问题：关于物理学与生物学的谈话》）。

1972年9月，为了庆祝Paul Dirac七十寿辰，在特里亚斯特（Tri-

este，意大利）举行了一个讨论会。会上提出的一些论文收入The Phys-ist’s Conception of Nature（ed.J.Mehra；Reidel 1973）（《物理学家心中的大自然》）。亦参见Heisenberg相同题目的一本书（1958）。

另一场庆祝活动是1979年的爱因斯坦诞辰百年纪念会，会上提出了若干如何看大自然的论文。例如，H.Woolf编辑的Some Strangeness in the Proportion：A Centennial Symposium to Celebrate the Achievements of Albert Einstein（Addison-Wesley 1980）（《比率中的某种异常：庆祝爱因斯坦的伟大成就百年纪念讨论会》）一书就收有好几篇文章讨论爱因斯坦对大自然的看法（爱因斯坦的同行们并非总是跟他看法一致），这些文章有的是涉猎广泛、探讨概念的，有的则提供了有关爱氏观点的有价值的历史材料。

非专业人员若想大致了解物理学家对大自然的看法，一本非常好的书是Richard Feynman根据自己的广播讲演编成的The Character of Physical Law（B.B.C.Publications 1965）（《物理定律的本性》）。

并非所有的人都认为物理学家们看得对。对某些早期思想的批评性评价，见L.S.Stebbing的Philosophy and the Physicists（Pelican 1944）（《哲学与物理学家》）。

对John Conway的“生命”游戏的描述，见Winning Ways（Academic Press 1982，VolII）（《得胜的方法》）。这本书是E.R.Berleka mp,J.H.Conway和R.K.Guy的著作。

# 注释

第一章

①H·邦迪，“宗教是一件好事”，《说谎的真理》（H.Bondi，'Religion is a good thing'，Lying Truths，编者：R.Duncan and M.Weston-Smith；Pergamon 1979版）。

②同上。

③K·佩德勒，《精神高于物质》（K.Pedlar,Mind over Matter,Thames Methuen 1981版），p.11。

④H·莫洛维茨，“重新发现精神”，《精神是我》（H.Morowitz，‘Rediscovering the mind'，The Mind's I，编者：D.R.Hof-stadter and D.C.Dennett；Harvester/Basic Books 1981版）。

第二章

①I·康德，《纯粹理性批判》（I.Kant,Critique of Pure Reason，编者，J.M.D.Meiklejohn；Dent 1934，1945版；初版1781）。

②教皇庇护十二世演讲的主要部分用英文刊载于《原子物理学家通讯》（Bulletin of the Atomic Scientists 8，pp.143-6，165（1952））。

③E·麦克穆林，“宇宙论与神学为何相联系？”，《二十世纪科学

与技术》（E.McMullin，‘How should cosmology relate to theology？'，The Sciences and Theology in the Twentieth Century，编者：A.R.Peacocke；Oriel 1981版）。

第三章

①塞缪尔·克拉克（Samuel Clarke），他在1704年所作的一组讲演中发挥了自己的宇宙学论，发表时书名为《上帝的存在和属性的论证》（A Demonstration of the Being and Attributes of God）。后来加上1905年所作的又一组讲演，发表时书名为《论上帝的存在和属性、论自然宗教不可改变的义务、基督教启示中的真理与肯定性》（A Discourse Concerning the Being and Attributes of God,the Obligations of Natural Religion,and theTruth and Certainty of the Christian Rrevelation,John and Daul Knapton；London，1738，第9版）。

②阿奎那，《神学大全》（Aquinas,Summa Theologiae，编者T.Gilby；Eyre&Spottiswoode 1964版）。

③克拉克（Clark），见所引著作，页12～13。

④D·休谟，《关于自然宗教的对话》（D.Hume,Dialogues Concerning Natural Religion，编者H.D.Aiken Hafner 1969版；1779初版），第IV部分。

⑤希波的圣奥古斯丁，“关于时间的起始”，《论上帝之城》（St. ；Augustine of Hippo，‘On the beginning of time'，The City of God，译者：M.DOds；Hafner 1948版）。

⑥J·A·惠勒，“创世与观察者”，《几个特殊学科中的基本问题》（J·A·Wheeler，‘Genesis and observership’，Foundational Problems in the Special Sciences，编者R.E.Butts and K.J.Hintikka；Reidel

1977版）。

⑦同上。

⑧J·A·惠勒，“黑洞之外”，《比例中的几个奇异现象》（J.A Wheeler，‘Beyond the black hole'，Some Strangeness in the Pro-portion，编者：H.Woolf；Addison-Wesley 1980版）。亦见“物理学受宇宙论规律的支配吗？”，《牛津论丛·量子引力》（‘Is physics legislated by cosmology'，Quantum Gravity：An Oxford Symposium，编者：C.J.Isham,R.Penrose and D.W.Scia-ma；Oxford：Clarendon Press 1975版，以及《时间边界》（Frontiers of Time,North-Holland 1979版）。

⑨J·R·哥特（J.R.Gott）在《自然》杂志[Nature 295，304（1982）]提出所谓“冒泡宇宙论”（‘bubble cosmology'），并在《科学》（The Sciences，1982）一书中较为通俗地作了阐释。类似的思想亦见于佐藤克彦等著《理论物理学的进展（通信集）》-Progress in Theoretical Physics（Letters））65，1443，1981）]。

⑩R·斯温伯恩，《论上帝的存在》（R.Swinburne,The Existence of God,Oxford：Clarendon Press 1979版），p.122。

第四章

①斯温伯恩，见所引著作，第七章。

②同上，页131～132。

③E·W·巴尔尼斯，《科学理论与宗教》（E.W.Barnes,Scien-ti fic Theory and Religion,Cambridge University Press 1933版），p.595。

④P·C·W·戴维斯，《时间不对称物理学》（P.C.W.Davies,The Physics of Time Asymmetry,Surrey University Press/Uni-versity of California Press 1974版）。

⑤S·W·霍金，“引力崩塌中可预测性的崩溃”，《物理学杂志》[S.W.Hawking，‘Breakdown of predictability in gravitational collapse'，Physical Review D14，2460（1979）]；亦见《科学美国人》[ScientificAmerican 236，34（1977）]。

⑥同上。

第五章

①A·柯斯特勒，“仅仅是……？”，《说谎的真理》（A.Koestler，‘Nothing but……？’，Lying Truths，编者：R.Duncan and M.Weston-Smith；Pergamon 1979）。亦见《还原论之外生命科学的新透视》（Beyond Reductionism-New Perspectives in the Life Sciences，编者：A.Koestler and J.R.Smythies；Hutchinson 1969版），以及《雅努斯：总结》（Janus.A Sum-ming Up,Kowstler编；Vintage 1979版）。

②D·玻姆，《完整性和其中包含的有序》（D.Bohm,Wholeness and the hnplicate Order,Routledge&Kegan Paul 1980版）。

③E·薛定谔，《生命是什么？》（E.Schr？dinger,What is Life？，Cambridge University Press 1946版），p.77。此书现已同薛氏另一名著《精神和物质》（Mind and Matter）合成一部，书名用的就是上述二书名的复合，CUP 1967版。

④同上，页76。

⑤F·克里克，《生命本身：起源和本质》F.Crick,Life Itself：Its

Origin and Nature,Macdonald/Simon&Schuster 1982版）。

⑥E·麦克穆林（E.McMullin），见所引著作，页47。

第六章

①《新凯斯里克百科全书》（New Catholic Encyclopedia,McGraw-Hill 1967版），vol.13，p.460。

②R·J·希斯特，《感觉的问题》（R.J.Hirst,The Problems of Percept ion,Allen&Unwin 1959版，p.181。笛卡儿的二元论范例是他在其主要著作《方法谈》[Le Discours de la Methode（1637）]和《哲学原理》-Principia Philosophiae（1644）]中提出的。见《笛卡儿哲学著作集》（The Philosophical Works of Descarte s，译者：E.S.Haldane and G.R.J.Ross；2 vols.，Cambridge University Press 1967版）。

③G·赖尔，《精神的概念》（G.Ryle,The Concept of Mind,Hutchinson 1949版，后多次重印）。

④同上，页20。

⑤《新凯斯里克百科全书》，见所引著作，页471。

⑥见所引著作，页23。

⑦《精神是我》，页6。

⑧D·R·霍夫斯塔特，《哥德尔、埃舍尔、巴赫》（D.R.Hof-stadter,Godel,Escher,Bach,Basic,Books，1979版，p.577。

⑨D·M·麦克埃，《发条形象》（D.M.MacKay,The Clockwork Image,Inter-Varsity Press 1974版），第九章。

⑩J·A·福德，“精神-肉体问题”，《科学美国人》（J.A.Fodor，‘The mind-body problem’，Scientific American[January 1981）]。

第七章

①T·里德，《论人的智力》（T.Reid,Essays on the Intellectual Powers of Man，编者：A.D.Woozley；MIT Press 1969版；初版1785），论文III，第四章。

②赖尔，见所引著作，页187。

③D·休谟，《人性论》（D.Hu me,A Treatise of Human Nature，编者：P.H.Nidditch；Oxford University Press 1978版；初版1739），第一册，第四部分，第六章。

④J·洛克，《人类理解论》，第27章（J.Locke,Essay Concerning Hu ma n Un de rs ta ndi ng，编者：A.D.Wo ozle y；D ent 197 6版；初版1690）。

⑤J·R·卢卡斯，“大脑、机器和哥德尔”，《大脑和机器》（J.R.Lucas，‘Minds,machines and Godel'，Minds and Machines，编者：A.R.Anderson；Prentice-Hall 1964版），p.57。

⑥A·J·艾尔，《哲学的中心问题》（A.J.Ayer,The Central Questions of Philosophy,Weidenfeld&Nicolson 1973版；Pen-guin 1977版），p.119。

⑦霍夫斯塔特，见所引著作，页697。

⑧卢卡斯，见所引著作。

⑨霍夫斯塔特，见所引著作，页709。

⑩见“我在哪里？”，《脑猝变》，作者：D·C·登奈（‘Where am I？'，Brainstorms byD.C.Dennett,Bradford Books 1978版）。

（11）麦克埃（MacKay），见所引著作，页75。

第八章

①D·玻姆，见所引著作，页134。

②E·威格纳“论精神肉体问题”，《沉思的科学家》（E.Wigner，‘Remarks on the mind-body Question'，The Scientist Speculates，编者：I.J.Good；Heinemann 1962版）。

③B·S·德威特，“关于量子力学的多宇宙解释”，《量子力学的基础》（B.S.DeWitt，‘The many-universes interpretation of quantum mechanics'，Foundations of Quantum Mechanics，编者：B.d'Espagnat；Academic Press 1971版）。

第九章

①见“麦克塔戈特、固定性和实现”，《时间，还原和现实》（‘McTaggart,fixity and coming true'，Time,Reduction and Reality，编者：Richard Healey；Cambridge University Press 1981版）。

②圣安赛尔姆，最早主张上帝超越时间的人有A·M·S·波伊提乌（A.M.s.B oethius,c.480～524）。见《哲学的慰藉》（The Consolation of Philosophy，编者W.Anderson,Centaur 1963），第5、6节。

③P·蒂里希，《系统神学》（P.Tillich,Systematic Theology[S.C.M.1978）]，vol.I,P.305。

④K·巴特，《教会教义学》（K.Barth,Church Dogmatics II（i），译者：C.W.Bromiley and T.F.T orranc e；T.&.T.Clark 1956版），p.620。

第十章

①麦克埃，见所引著作，页78。

第十一章

①J·A·惠勒著《引力论》（J.A.Wheeler in Gravitation，编者：C.W.Misner,K.S Thorne and J.A.Wheeler；Freeman 1973版），p.1197。

第十二章

①《威廉·佩利著作集》（The Works of William Paley,Oxford；Clarendon Press 1938版）。

②R·斯温伯恩，《上帝的存在》（R.Swinburne,The Existence of God,Oxford：Clarendon Press 1979版）

③B·卡特，“大数重合与宇宙论中的人存原理”，《宇宙论诸理论与观察结果的冲突》（B.Carter，‘Large number coincidences and anthropic principle in cosmology’，Confrontation of Cosmological Theories with Observation，编者：M.s.Longair；Reide 1974版）。这篇论文基于卡特此前的一篇更为广泛的探讨，题目为“自然中数重合的重要意义”（‘The significance of numerical coin-cidences in nature’，但该文从未全文发表过）。

第十三章

①R·彭罗斯，“奇点与时间不对称性”，《广义相对论：爱因斯坦百年巡礼》（R.Penrose，‘Singularities and time-asymmetry'，General Relativity：An Einstein Centenary Survey，编者：S.；W.Hawking and W.Israel；Cambridge University Press 1979版）。

②卡特，见所引著作。

第十五章

①E·R·哈里森，《宇宙学》（E.R.Harrison,Cosmology,Cambridge University Press 1981版），p.360。

②N·尼科尔森，“扩张中的宇宙”，（N.Nieholson，‘The expanding universe'in The Pot Geranium,Faber&Faber 1954版）。

第十六章

①A·H·古斯，“关于物质的起源、能和宇宙的熵的思考”，见《物理学的渐近领域：弗兰西斯·洛纪念文集》（A.H.Guth，‘Speculations on the origin of the matter,energy and entropy of the universe'in Asymptotic Realms of Physics：A Festschri ft in Honor of Francis Low，编者：A.H.Guth,K.Huang and R.L.Jaffe；MIT Press 1983版）。

第十七章

①P·A·M·狄拉克，“物理学家对自然的描述的发展史”，《科学美国人》[P.A.M.Dirac，‘The evolution of the physicist's picture of nature'，Scientific American（May 1963）]。

②D·玻姆，见《一个物理学问题：物理学和生物学中的对话》（D.Bohm,in A Question of Physics：Conversations in Physics and Biology，编者：P.Buckley and F.D.Peat；Routledge&Kegan Paul 1979版），p.129。

③A·爱因斯坦，《科学论文集》（A.Einstein,Essays in Science,Philosophical Library,New York 1934版）。

④J·A·惠勒，见所引巴克利与F·D·皮特的著作，页60（J.A.Wheeler in Buckley and Peat）。

⑤D·Z·弗里德曼和P·范·纽温惠仁，“超引力与物理学定律的统一”，《科学美国人》[D.Z.Freedman and P.Van Nieuwen-huizen，‘Supergravity and the unification of the laws of physics'，Scientific American（February 1978）]。

⑥R·培根，《大著作》（R.Bacon Opus Majus，译者：Robert Belle Burke；University of Pennsylvania Press 1928版）。

⑦J·A·惠勒，见《引力论》（编者：C.W.Misner,K.S Thorne and J.A.Wheeler；Freman 1973版），p.1212。

⑧A·爱因斯坦致助手恩斯特·施特劳斯，引自《爱因斯坦百年纪念文集》（A.Einstein to his assistant Ernst Straus,as quoted in Einstein：A Centenary Volume，编者：A.P.French；Heine-mann 1979版），p.128。

⑨R·P·费恩曼，《物理定律的本性》（R.P.Feynman,The Character of Physical Law,B.B.C.Publication 1965版），pp.124～5。

⑩F·J·戴森，“宇宙中的能”，《科学美国人》

[F.J.Dyson，‘Energy in the universe’，Scientific American（September 1971）]。

# 名词对照表

A

阿奎那，圣托马斯Aquinas,St.Thomas，

阿斯贝，阿莱纳Aspect,Alaine

阿西莫夫，艾萨克Asimov,Isaac

爱因斯坦，阿尔伯特Einstein,Albert

埃弗列特，休Everett,Hugh

埃弗列特对量子力学的解释Everett interpretation of quantum mechanics

艾丁顿，亚瑟爵士Eddington,Sir Arthur，

埃京，曼弗雷德Eigen,Manfred

埃庇米尼底斯的悖论Epeminides'paradox

埃舍尔Eseher.M.C.

艾耶尔，A.J.Ayer,A.J.

安德森，卡尔Anderson,Carl，

暗含的秩序Implicate order

安赛尔姆，圣Anselm,St.

奥古斯丁，圣Augustine,St.

奥科姆剃刀原理Oceam’s razor

B

八正道Eight foldway

巴尔尼斯，E.W.Barnes,E.W.

巴罗，约翰Barrow,John，

巴斯，卡尔Barth,Karl.

鲍依修斯Boethius

暴涨（宇宙中的）Inflation（in the universe）

贝尔，约翰Bell,John

贝尔的不等式Bell's inequality

贝肯斯坦，雅各布Bekenstein,Job

贝罗索夫-扎布汀斯基反应Belousov-Zhabotinski reaction

本能Instinct

玻尔，尼尔斯Bohr,Niels

玻尔-爱因斯坦之争论Bohr-Einstein controversy

波-粒二象性Wave-particle duality

玻姆，大卫Bohm,David

柏拉图Plato

不可预测性Unpredictability

不明飞行物U.K.O.

不确定性原理Uncertainty principle

C

测量（量子论中的）Measurement（in quantum theory）

禅Zen

场，的概念Field,concept of

超自然的现象Paranormal phenomena

程序（计算机）Program（computer）

创世纪Book of Genesis

创世论（参看宇宙）Creationism

创生Creation

D

大爆炸（亦见宇宙的创生）Big bang

大崩塌Big crunch

大统一理论Grand unified theories

达尔文，查尔斯Darwin,Charles

达芬奇，列奥纳多da Vinci,Leonardo

达兰贝尔，让d'Alembert,Jean

戴森，弗里曼Dyson,Freeman

氘Deuterium

德威特，布莱希Dewitt,Bryce

德夏尔丹，戴拉de Chardin,Teilard

笛卡儿，勒内Descartes,Rene

狄拉克，保罗Dirac,Paul

蒂里希，保罗Tillich,Paul

电磁Electromagnetism

定律（规律），大自然的Laws,of nature

电子Eletron

电子对Positronium

动物电流Animal electricity

多神论Polytheism

多世界理论（见量子论）Many-worlds theory

E

恶（邪恶）Evil

二元论Dualism

F

反馈Feedback

反物质Antimatter

放射性Radioactivity

分子，生物的Molecules,biological

费恩伯格，杰拉尔德Feinberg,Gerald

费恩曼，理查德Feynman,Richard

费恩曼图Feyman diagrams

非决定论Indeterminism

冯·纽曼，约翰von Neumann,John

冯·魏泽克，卡尔von Weiszacker,Carl

G

盖尔曼，莫里Gell-Mann,Murray

盖勒，尤里Geller,Uri

概率（几率）Probability

感觉（感知）Perception（Sensation）

干扰（量子波的）Interference（of quantum waves）

高能物理High-energy physics

哥白尼，N.Copernicus,N.

哥德尔，库尔特Godel,Kurt

哥德尔定理Godel's theorem

戈尔德Gold,Thomas

功能主义Functionalism

古斯，艾伦Guth,Alan

观察Observation

观察者Observers

光Light

光子Photon

过去和将来Past and future

H

哈勃，埃德温Hubble,Edwin

哈里森，爱德华Harrison,Edward

氦Helium

海森伯，沃纳Heisenberg,Werner

海森堡的不确定性原理Heisenberg's uncertainty priniple

核力Nuclear force

赫斯特，R.J.Hirst,R.J.

黑矮星Black dwarf star

黑洞Black holes

黑尔斯，斯蒂芬Hales,Stephen

红移Redshift

还原论Reductionism，

惠勒，约翰Wheeler,John

混沌，原初的Chaos,primeval

霍夫斯塔特，道格拉斯Hofstadter,Douglas

霍金，史蒂芬Hawking,Stephen

霍伊尔，弗雷德爵士Hoyle,Sir Fred

活力论Vitalism

J

基本常数Fundamental constants

基本粒子Elementary particles

基督，耶稣Christ,Jesus

奇点（黑洞中的）Singularities,in black holes

吉恩斯，詹姆斯爵士Jeans,Sir James

机器（会思想的）Machines that think

技术Technology

计算机Computers

记忆Memory

加尔凡尼，卢奇Galvni,Luigi

加速器Accelerators

伽利略Galileo Galiei

将来与过去Future and past

教会Church

胶子Gluons

介子Mesons

进化论Evolution,theory of

金属弯曲Metal bending

精神Mind

的出现appearance of

离体的disembodied.

的二元论dualist theory of

上帝God

整体方面bolistic aspect

无限的infinite

的死后存在survival of

量子物理学中的in quantum physics

精神-肉体相互作用Mind-body interaction

精神-肉体问题Mind-body problem

精神作用于物质Mind-over-matter

决定论Determinism

K

卡普拉，弗里提约夫Capra,Fritjoy

卡特，布兰东Carter,Brandon

康德，伊曼纽尔Kant,Immanuel

康威，约翰Conway,John

克拉克，阿瑟Clarke,ArthurC.

克拉克，塞缪尔Clarke,Samuel

克里克，弗兰西斯Crick,Francis

柯普莱斯顿，F.C.，Co pleston,F.C.

柯斯特勒，阿瑟Koestler,Authur

科学Science

科学的真理Truth in science

可预测性Predictability

空间，的创生Space,creation of

有弹性的，有伸缩性的elastic

膨胀的expanding

的存在existence of

有限的finite

更多维数的higher dimensional

无限的infinite

与时间关联的linked to time

的量子特性quantum featuresof

作为合成的结构的as synthetic structure

空间弯曲Spacewarps

夸克Quarks

L

拉普拉斯，彼埃尔Laplace,Pierre

拉普拉斯计算Laplacian calculator

莱布尼茨，G.Leibniz,G.

力（大自然的）Fcrces（of nature）

电磁electromagnetic

的量子描述quan tum desc ript ion of

的统一unification of

核nuclear

强（作用）strong

弱（作用）weak

引力gravitational

意念力psychic forces

里德，托马斯Reid,Thomas

粒子（见基本粒子）Particles

量子论Quantum theory

测量行为act of measurement

波尔的解释Bohr's interpretation

和因果关系and causation

波的崩溃collapse of wave

和宇宙的创生and creation of the universe

爱因斯坦的质疑Einstein's objections to

引力的of gravity

和整体论and holism

和直觉and intuition

多宇宙（埃弗列特的解释）many-universes（Everett）interpretation

和精神and mind

实际应用practical applications

和可预测性and predictability

和实在and reality

和不确定性and uncertainty

和真空and the vacuum

量子引力（见量子论）Quantum gravity

量子宇宙学Quantum cosmology

灵（或幽灵）Ghost

灵魂（参看精神）Soul

圣经里提到的biblical reference to

的定义definition of

的处所location of

的死后继续存在survival of

和时间and time

灵体Aetheric body

鲁尔底斯Lourdes

卢卡斯，J.R.Lucas,J.R.

卢瑟福Rutherford

逻辑Logic

逻辑实证主义（见实证主义）Logical positivism

洛克，约翰Locke,John

罗素，贝特兰Russell,Bertrand

罗素的悖论Russell's paradox

罗伊尔，吉尔伯特Ryle,Gilbert

M

麦克埃，唐纳德MacKay,Donald

麦克穆林，厄南McMullin,Ernan

麦克斯韦，詹姆斯·克拉克Maxwell,James Clerk

麦氏妖Maxwell's demon

麦克塔戈特，J.McTaggart,J.

美，大自然规律中的Beauty,in the laws of nature

米勒，斯坦利Miller,Stanley

米勒-尤里实验Miller-Urey experiment

米勒尼，E.A.Milne,E.A.

秘学Occult

默顿，罗伯特Merton,Robert

莫洛维茨，哈罗德Moro witz,Harold

末世学Eschatology

魔法Magic

莫彼乌斯带M？bius band

目的论Teleology

N

能级Energy levels

能量Energy

尼科尔森，诺曼Nicholson,Norman

尼曼，尤瓦尔Ne'eman,Yuval

牛顿，艾萨克爵士Newton,Sir Isaac

牛顿力学Newtonian mechanics

P

派德勒，基特Pedlar,Kit

彭罗斯，罗杰Penrose,Roger

培根，罗杰Bacon,Roger

佩利，威廉Paley,William

庞卡莱Poincaré，Henri

庞卡莱循环Poincaré cycle

庇护十二世，教皇PlusXII，Pope

平方反比定律I nver se s quare law

普遍的精神Universal mind

普朗克，麦克斯Planek,Max

普朗克时Planck time

普里高津，以利亚Prigogine,Ilya

Q

奇迹Miracles

前几何Pregeometry

强子Hadrons

氢Hydrogen

轻子Leptons

邱，杰弗里Chew,Geoffrey

R

热，原初宇宙的Heat,of primeval universe

热力学，黑洞的Thermodynamics,of black holes

热力学第二定律Second law of thermodynamics

热力学平衡Thermodynamic equilibrium

仁慈（上帝的）Benevolence（of God）

人工智能Artificial intelligence

人类Man

人存原理Anthropic principle

认知科学Cognitive sciences

软件Software

弱作用力（见力）Weak force

S

萨拉姆，阿布杜斯Salam,Abdus

色动力学Chromodynamics

熵Entropy

和无序and disorder

和概率and probability

和生命and life

和信息and information

黑洞的of black holes

引力的gravitational

宇宙的cosmic

的增加increase of

最大的maximum

上帝God

神秘主义Mysticism

生命Life

生命力Life-force

圣安赛尔姆Saint Anselm

圣经Bible

设计Design（in the universe）

时间Time

绝对的absolute

之矢arrow of

不对称asymmetry of

的开始beginning of

与因果关系and causation

的创生creation of

的膨胀dilation of

的终结end of

有弹性的，有伸缩性的elatic

的存在existence of

有限的finite

之流逝flow

和引力and gravity

无限的infinite

的感知perception of

的倒转reversal of

作为合成的结构as a synthetic structure

旅行travel

普遍的universal

时间流Time-flux

时间弯曲Timewarps

事件Events

第一first

的时间顺序time order of

视界（空间的）Horizon（in space）

世界World

精神的mental

物质的physical

时空，的界限（边界）Spacetime,boundary of

实在和量子论Reality and quantum theory

实证主义Positivism

数学，无限的Mathematics,of infinity

双生子效应Twins effect

斯济提拉，列奥Leo Szilard

思想Thoughts

思想机Thinking machines

宿命论Fatalism

速子Tachyons

随机性Randemness

T

汤川秀树Yukawa Hideki

天，天堂Heaven

同位旋Isotopic spin

同一性（人的）Identity（personal）

统一理论（力的）Unified theories（of forces）

图灵，艾伦Turing,Alan

脱氧核糖核酸D.N.A.

W

外来生命，天外生命Alien life

威格纳，尤金Wigner,Eugene

唯物主义Materialism

唯心主义Idealism

温伯格，斯蒂芬Weinberg,Steven

温伯格-萨拉姆理论Weinberg-Salam theory

物理（学）Physics

物质，的建筑砌块Matter,building blocks

无限，无穷Infinity

无穷（量子场论中的）Infinities（in quantum field theory）

无知原理Ignorance principle of

X

细胞核Nucleus of cell

细菌Bacterium

夏皮罗，罗伯特Shapiro,R.

显现性Emergent qualities

现在的概念Now,concept of

先知先觉Precognition

相对论Relativity,theory of

广义general

狭义special

信使粒子Messenger particles

信息Information

信仰疗法Faith healing

行为主义Behaviourism

休谟，大卫Hume,David

悬浮Levitation

旋转（亚原子粒子的）Spin（of subatomic particles）

薛定谔，埃尔温Schr?dinger,Erwin

薛定谔猫Schr?dinger's cat

Y

亚里士多德Aristotle

延迟选择实验Delayed choice experiment

杨，托马斯Young,Thomas

杨的实验Young's experiement

耶稣基督Jesus Christ

意念致动Phychokinesis

意识Consciousness

意志Volition

因果（原因和结果）Cause and effect

逆动的retroactive

的颠倒reversal of

引力Gravity

引力崩塌Gravitational collapse

引力场Gravitational field

引力子Graviton

硬件Hardware

永生Immortality

尤里，哈罗德Urey,Harold

宇宙，年龄Universe,age of

宇宙热辐射（来自大爆炸的）Cosmic heat radiation（from big bang）

宇宙学Cosmology

原初汤Primeval soup

原初的宇宙（见宇宙）Primeval universe

无素，的合成Elements,synthesis

原子，的构造Atoms,structure of

原子核Nuclei of atoms

约瑟夫，圣，库帕第诺的Joseph of Cupertino

Z

再世Reincarnation

造物主别针类比Creator's pin analogy

赵忠尧Chao,C.Y.

振荡宇宙Oscillating universe

真空，的量子性质Vacuum,quantum nature of

整体论Holism

正电子Positron

质量Mass

质子Proton

秩序（有序），复杂的Order,of complexity

钟表，的行为Clocks,behaviour of

中微子Neutrinos

中子Neutron

重子Baryons

自创造的宇宙Self-creating universe

自发产生（生命的）Spontaneous generation（of life）

自复制的机器Self-reproducing machines

自我Self,the

自由意志Free will

自指Self-reference

自组织Self-organization

祖卡夫，盖里Zukav,Gary

W粒子W particle

Z粒子Z particle

μ子Muons

π子Pions

τ子Taus

# 译后记

《上帝与新物理学》的作者保罗·戴维斯，英国物理学家，University of Newcastle upon Tyne的理论物理学教授，主要研究现代宇宙学，写过好几本很受欢迎的有关宇宙学的普及读物。

《上帝与新物理学》一书，涉及很多宇宙学方面的问题，但主要谈的是现代物理学的进展对宗教教义，尤其是对西方文明的基本元素之一的基督教教义的影响。作者自云："本书不是一本谈宗教的书。本书谈的是新物理学对以前属于宗教的问题产生的影响。"

作者在本书的开始，引用了爱因斯坦的两句话："没有科学的宗教是盲目的。没有宗教的科学是跛足的。"这两句话，大约也是代表了戴氏本人的意见。所谓"没有科学的宗教"在我们或许难以理解。它显然是与"有科学的宗教"相对而言的。对于"有科学的宗教"（如现今西方的基督教），人们可以进行自由的探讨、批判。而"没有科学的宗教"，你只能老老实实地信，不能问，不能批判。一种信仰，若只能信，不能问也不能批判，就只能是一种盲目的信仰，当然也就是盲目的宗教。人类受过和正在受着的盲目宗教的害太多了。那么，"没有宗教的科学"又是指什么呢？这里所说的"宗教"还是指信仰。常听有人说，"科学知识的真正源泉是观察"。这话当然不错，确实有一些科学知识起源于观察。但这话只能说是部分正确，因为还有一些科学知识起源于猜想。例如，早在几千年前人类就猜想大地是球形的。那时谁也没能真正"观察"到地球是一个"球"。要证实这一猜想，从而使这猜想成为科学知识，就得进行一番科学实验。而若想进行这样的科学实验，就得有坚定如宗教的信仰。于是，宗教与科学就这样联系了起

来。假如当年的哥伦布、麦哲伦不是坚信大地是一个球体，他们就想不到去进行环球航行。反观中国七下西洋的郑和，虽然率领的船只、船队堪称当时的世界先进水平，但就增进人类在地理方面的科学知识而言，成就却相形见绌。就郑和来说，他根本就没有想到大地是否是球形的问题，因而也就没有什么信不信的问题。这也就是为什么说“没有宗教的科学是跛足的。”自然，我们还是应该为郑和的壮举自豪，因为跛足的科学仍是科学，但我们应该承认跛足毕竟不如不跛足。中国的科学从近代开始落后了，其原因很多。有人认为，其中一个原因是中国从事科学研究的人过于世俗，太缺乏宗教精神，因而虽然心灵手巧，很有技能，却难有重大的科学发现。这种看法是否有道理，有待于科学史家们详细论证。

本书除“前言”外，共十七章，结合宗教、哲学讲述了最新的科学理论和发现。在本书的“前言”中，作者说：本书的主题是我所谓的四大存在问题：

为什么大自然的规律是现在这样的？

为什么宇宙是由现在组成它的各种东西所组成的？

这些东西是如何起始的？

宇宙如何获得了组织？

这四大问题以前是宗教的、玄学的问题，但近年来，“科学已进展到可以认真地解决以前是属于宗教的问题的地步了”，这是“严肃的宗教思想者不能忽视”的。

在接下来的十七章里，作者首先集中讲述了他对科学和宗教的看法，然后以流畅优美、引人入胜的文笔，介绍了新物理学对宇宙的创生和终结、时间、生命的本质、灵魂、自我、自由意志与决定论等一系列问题的研究进展。作者身为科学家，当然对科学的进步津津乐

道，但也以科学的精神批判了科学，指出“科学对工业社会的影响是好坏参半”。对盲目的宗教，作者予以无情的谴责，指出，“尽管大多数宗教都赞美爱心、和睦、谦卑，并将这些称之为美德，但世界上各大宗教组织的历史却常常是以仇恨、战争、傲慢为其特色的。”然而。作者并没有完全否定宗教。在本书结尾时，作者在指出科学并不能解决所有问题之后写道：“我深信，只有从各个方面全方位地了解世界，从还原论和整体论的角度，从数学和诗的角度，通过各种力、场、粒子，通过善与恶，全方位地了解世界，我们才能最终了解我们自己，了解我们的家——宇宙背后的意义。”显然，作者在这里把关注善恶的宗教也看作是我们认识世界、认识自己的途径之一。

本书虽是一本普及性的书，但“包含一些具有真正学术价值的材料”，涉及很多奋战在科研前沿的科学家正在思考、解决的问题，所以也值得有关专业人员一读。

应该说明的是，译者学力不足，翻译中多有错误，承蒙一位科学家将译稿校读一遍，修补了不少错漏。在此谨表示衷心感谢。译文中若仍有错误，应由译者本人负责，欢迎读者指正。译者在此也应衷心感谢山东大学外文系的李绍明君。李君拿出自己宝贵的时间帮助誊抄了部分译稿，并协助重新编排了原书的索引，解决了很多问题。

徐培

1991年于北京气象学院

[1]《圣经》记载，上帝通过摩西向以色列人传达的十诫中，有“不可贪恋人的牛驴”的话，见《圣经·出埃及记》第20章第17节。——译注

[2]Creation，在英文中有“创造”，“创作”，“发生”之类的意思。英文圣经中，上帝造天地、造人，用的一般都是Creation或其同源词，因而这个词很有宗教意味。所以，本书的作者要在这里专门澄清Creation的意思。——译注

[3]以色列人，自称神人下凡，曾多次举行多种特异功能的表演，在世界上引起广泛的轰动。后来查明，他的那些特异功能是巧妙的魔术。——译注

# 物理学的困惑

# 总序

科学，特别是自然科学，最重要的目标之一，就是追寻科学本身的原动力，或曰追寻其第一推动。同时，科学的这种追求精神本身，又成为社会发展和人类进步的一种最基本的推动。

科学总是寻求发现和了解客观世界的新现象，研究和掌握新规律，总是在不懈地追求真理。科学是认真的、严谨的、实事求是的，同时，科学又是创造的。科学的最基本态度之一就是疑问，科学的最基本精神之一就是批判。

的确，科学活动，特别是自然科学活动，比较起其他的人类活动来，其最基本特征就是不断进步。哪怕在其他方面倒退的时候，科学却总是进步着，即使是缓慢而艰难地进步，这表明，自然科学活动中包含着人类的最进步因素。

正是在这个意义上，科学堪称为人类进步的“第一推动”。

科学教育，特别是自然科学的教育，是提高人们素质的重要因素，是现代教育的一个核心。科学教育不仅使人获得生活和工作所需的知识和技能，更重要的是使人获得科学思想、科学精神、科学态度以及科学方法的熏陶和培养，使人获得非生物本能的智慧，获得非与生俱来的灵魂。可以这样说，没有科学的“教育”，只是培养信仰，而不是教育。没有受过科学教育的人，只能称为受过训练，而非受过教育。

正是在这个意义上，科学堪称为使人进化为现代人的“第一推

动”。

近百年来，无数仁人志士意识到，强国富民再造中国离不开科学技术，他们为摆脱愚昧与无知作了艰苦卓绝的奋斗。中国的科学先贤们代代相传，不遗余力地为中国的进步献身于科学启蒙运动，以图完成国人的强国梦。然而应该说，这个目标远未达到。今日的中国需要新的科学启蒙，需要现代科学教育。只有全社会的人具备较高的科学素质，以科学的精神和思想、科学的态度和方法作为探讨和解决各类问题的共同基础和出发点，社会才能更好地向前发展和进步。因此，中国的进步离不开科学，是毋庸置疑的。

正是在这个意义上，似乎可以说，科学已被公认是中国进步所必不可少的推动。

然而，这并不意味着，科学的精神也同样地被公认和接受。虽然，科学已渗透到社会的各个领域和层面，科学的价值和地位也更高了，但是毋庸讳言，在一定的范围内，或某些特定时候，人们只是承认“科学是有用的”，只停留在对科学所带来的后果的接受和承认，而不是对科学的原动力、科学的精神的接受和承认。此种现象的存在也是不能忽视的。

科学的精神之一，是它自身就是自身的“第一推动”。也就是说，科学活动在原则上是不隶属于服务于神学的，不隶属于服务于儒学的，科学活动在原则上也不隶属于服务于任何哲学。科学是超越宗教差别的，超越民族差别的，超越党派差别的，超越文化的地域的差别的，科学是普适的、独立的，它自身就是自身的主宰。

湖南科学技术出版社精选了一批关于科学思想和科学精神的世界名著，请有关学者译成中文出版，其目的就是为了传播科学的精神，科学的思想，特别是自然科学的精神和思想，从而起到倡导科学精神，推动科技发展，对全民进行新的科学启蒙和科学教育的作用，为

中国的进步作一点推动。丛书定名为《第一推动》，当然并非说其中每一册都是第一推动，但是可以肯定，蕴涵在每一册中的科学的内容、观点、思想和精神，都会使你或多或少地更接近第一推动，或多或少地发现，自身如何成为自身的主宰。

《第一推动丛书》编委会

# 绪言

上帝或诸神，可能有，也可能没有。不过确实有东西令我们对神圣的追寻更加高贵。另外，人们走过的那些将我们引向真理更深层次的每一条道路，说明还有东西令我们的追寻更为人性。有的人在沉思冥想或祷告中寻求卓越，另一些人则默默地为人类伙伴服务，还有些人很幸运，拥有超群的才华，通过艺术实践而达到巅峰。

走进生命最深层问题的另一途径就是科学。并非所有科学家都是探索者，多数都不是。但在每个科学分支领域，都有那样的科学家，渴望了解学科最基本的真理。数学家想知道数是什么，数学真理描述了什么；生物学家想知道生命是什么，生命是如何起源的；物理学家想知道空间和时间是什么，世界是怎么形成的。这些基本问题很难回答，也很少有直接的进步。只有很少的科学家能耐得住那样的寂寞。这是最冒险的事业，也有最大的回报：一旦有人回答了学科基础的某个问题，就将改变我们知道的一切。

科学家的使命是为我们的知识宝库不断添加新的东西，所以他们每天都面对着未知的事物。在学科基础领域工作的科学家们都明白，科学大厦的砖块决不像人们想象的那么坚固。

本书讲述的是在最深层次认识自然的故事。鼓吹它的是那些在努力延伸我们的物理学基本定律的科学家们。我要讲述的时期——大约从1975年开始——是我个人理论物理学生涯的几十年，大概也是自开普勒和伽利略400年前从事物理学以来最奇异、最令人沮丧的几十年。

我讲的故事在某些人读来可能像悲剧。老实说——说来好笑——我们失败了。我们继承的这门科学（物理学），长久以来在惊人地发展着，简直成了其他科学的楷模。在过去的两个多世纪里，我们极大扩展了对自然律的了解。但是今天，尽管我们付出了艰巨的努力，我们对那些定律的认识并不比20世纪70年代更多。

30年过去了，基础物理学却没有重大的进步，这是多么不同寻常的事情啊！即使我们回溯200年，当科学还是富家子弟的享乐时，也不曾有过这样的事情。至少在18世纪后期，大约每四分之一世纪都会出现关键问题的重大进步。

到1870年，当拉瓦锡（AntoineLavoisier）的定量化学实验证明物质守恒时，牛顿的运动和引力定律已经流行几乎100年了。虽然牛顿为我们提供了认识自然万物的框架，前景仍然广阔无边。那时，人们才开始认识物质、光、热的基本事实，正在揭开电磁等神秘现象的秘密。

在接下来的25年里，那些领域出现了重大发现。我们明白了光是一种波。我们发现了带电粒子间的力的规律。因为道尔顿（JohnDalton）的原子理论，我们对物质的认识向前飞跃了一大步。我们有了能量的概念，用光的波动理论解释了干涉和衍射，还探讨了电阻和电与磁之间的关系。

现代物理学的几个基本概念出现在1830年到1855年的四分之一世纪里。法拉第（MichaelFaraday）提出了场传递力的观点，极大促进了我们对电磁现象的认识。同在那个时期，能量守恒定律也随着热力学第二定律确立起来了。

接着的四分之一世纪，麦克斯韦（JamesClerkMaxwell）进一步发展了法拉第的场的思想，形成了我们今天的电磁学。麦克斯韦不仅统一了电和磁，还解释了光是一种电磁波。1867年，他用原子理论解

释了气体的行为。同时，克劳修斯（RudolfClau-sius）提出了熵的概念。

从1880年到1905年，我们发现了电子和X射线。热辐射的研究经历了几个阶段，最终导致普朗克（MaxPlanck）在1900年发现了描述辐射的热性质的正确公式——它点燃了量子革命的燎原烈火。

1905年，爱因斯坦26岁。虽然他过去关于热辐射的物理研究在后来被证明是科学的一大贡献，但他还是没能找到学术工作。不过那还只是热身，他很快就看准了一个基本的物理学问题：首先，如何才能让运动的相对性与麦克斯韦的电磁定律协调起来？他在狭义相对论里告诉了我们答案。我们应该把化学元素看做牛顿的原子吗？他证明确实应该那样。我们如何协调光的理论与原子的存在性呢？他也回答了，而且证明光既是波也是粒子。所有这些都发生在1905年，在他作为专利局技术员的空闲时间里。

爱因斯坦的思想爆发持续了四分之一世纪。到1930年，我们已经有了他的广义相对论，它革命性地宣称空间的几何不是固定的，而是随时间演化的。他在1905年揭示的波粒二象性成长为丰满的量子理论，使我们具体认识了原子、化学、物质和辐射。同在1930年，我们还认识了宇宙包含着大量像银河系一样的星系，而且它们在相互离开。这个现象的意义尚不清楚，但我们知道我们生活在一个膨胀的宇宙中。

随着量子理论和广义相对论成为我们对世界的认识的一部分，20世纪物理学革命的第一幕结束了。许多物理教授对各自专业领域的革命感到不安，他们宽慰自己，希望还能按照常规的方式做科学，而不必关心那些基本假设。可惜他们高兴得太早了。

爱因斯坦是在下一个四分之一世纪的最后那年（1955年）去世的。那时我们学会了怎样和谐地融合量子理论与狭义相对论，这是戴

森（FreemanDyson）和费曼（RichardFeynman）那一代人的伟大成就。我们发现了中子、中微子和千百种其他看起来基本的粒子。我们还认识到千变万化的大自然由四种力主宰：电磁力、引力、强核力（将原子核束缚在一起的力）和弱核力（决定衰变的力）。

再过四分之一世纪就到了1980年。我们那时构造了一个能解释所有基本粒子和力的实验结果的理论——即所谓的基本粒子的标准模型。例如，标准模型精确告诉我们质子和中子如何由夸克构成，夸克又如何通过胶子（强核力的传递者）而束缚在一起。在基本物理学历史上，我们第一次看到理论赶上了实验。从那时以来，还没有一个实验与标准模型或广义相对论相矛盾的。

我们的物理知识从微观走向宏观，现在走进了一门新的宇宙学——大爆炸理论已经成为常识。我们发现我们的宇宙不仅有恒星和星系，还有像中子星、类星体、超新星和黑洞那样的奇异天体。到1980年，霍金大胆预言了黑洞辐射。天文学家们也有证据证明宇宙包含着大量暗物质——也就是既不发光也不反射光的某种形式的物质。

1981年，宇宙学家古斯（AlanGuth）提出一幅“暴胀”图景来描述宇宙的早期历史。大致说来，他的理论认为宇宙在极早时期经历了急剧扩张的一幕，这就解释了为什么宇宙在各个方向是那么相同。暴胀理论的预言看起来很可疑，不过10年前开始出现了倾向它的证据。到我写这本书的时候，还存在几个疑难，不过总的证据还是支持暴胀预言的。

于是，到1981年时，物理学已经历了200年的茁壮成长。一个接着一个的发现深化了我们对自然的理解，因为理论和实验在每个时刻都手牵手地前进着。新的思想被检验和证实，新的实验发现得到了理论的解释。可是接下来，20世纪80年代之初，物理学的脚步停了。

我是粒子物理学标准模型建立后成长起来的第一代物理学家之

一。我与大学和研究生院的朋友们见面时，大家经常问“我们发现了什么值得我们这一代人骄傲的东西吗？”如果要说新的基本发现——被实验确立并由理论解释了的发现，像上面说的那些发现——那么我们只好承认，“没有！”外斯（MarkWise）是标准模型之外的粒子物理学的顶尖理论家。最近，在我工作的加拿大安大略省沃特卢圆周（Perimeter）理论物理研究所的一次研讨会上，他谈了基本粒子质量来源的问题。“我们在这个问题上败得很惨，”他说，“如果要我现在谈费米子的质量问题，我可能谈到20世纪80年代就无话可说了。”[1]接着他给我讲了一个故事：1983年，他和约翰?普雷斯基尔（JohnPreskill，也是一流的理论家）去加州理工学院任教。“约翰和我坐在他的办公室里聊天……你知道，物理学的大佬们曾经都在加州理工，而我们现在也来了！约翰说，‘我不会忘什么是要紧的事情。’于是他选择了夸克和轻子的质量，他把问题写在一张黄纸片上，然后把它贴在公告牌上……这样就不会忘记要为它们而工作。15年后，我走进他的办公室……谈点儿事情，我看了看公告牌，那纸片还在呢，但阳光已经洗净了上面的文字。所以问题也就消失了！”

平心而论，我们在过去的几十年里还是有两个实验发现：中微子有质量，宇宙的主角似乎是某种神秘的令膨胀加速的暗能量。但我们还不知道为什么中微子（或其他任何粒子）有质量，也不能解释其质量的数值。至于暗能量，现有的理论还不能解释。暗能量的发现不能算一个成果，因为它说明我们都忽略了某个重要的事实。除了暗能量外，我们再没发现什么新粒子和新的基本作用力，也没遇见过去25年所不曾知道的新现象。

不过也别误会。我们在过去25年当然也是忙忙碌碌的，成功地把确立的理论应用于不同的对象：材料的性质、生物的分子物理学、巨大星团的动力学。至于我们对自然律的认识的扩展，确实没有实在的进步。我们探索过很多优美的思想，也做过令人注目的粒子加速实验和宇宙观测，但它们主要是为了证实现有的理论，几乎没有什么飞

跃，当然也没有像过去200年那么确定或重要的发现。如果这样的事情发生在运动场或商场，那就是撞墙了，走霉运了。

物理学为什么突然陷入了困境？我们能为它做些什么？这就是本书的中心问题。

我是个乐天派，很长时间都不愿承认我自己经历的这个物理学时期会是那么沉寂。我和许多朋友一样，满怀希望地走进科学，期待着能为那个飞速发展的领域做出重大贡献，结果，我们却必须面对一个令人震惊的事实：我们不像我们的前辈，没有发现任何能流传后代的东西。这令很多人产生了危机；而更重要的是，它还带来了物理学的危机。

在过去的30年，理论粒子物理学的主要挑战是更深入地解释标准模型。这方面做了很多事情。我们提出了一些新理论，还进行过很详尽的探索，但都没能得到实验的证实。问题的症结在于，在科学中，一个理论要令人信服，它必须为尚未进行的实验做出新的预言——不同于从前理论的预言。而一个实验要有意义，它必须有可能产生与那个预言不一致的结果。如果实验结果和预言不一致，我们就说理论被证伪了——即它很可能被证明是错的。理论还必须是可以证实的，它应该能证明只有它才有的新预言。只有当理论经过了检验而结果与理论一致时，我们才能把它提升到那些正确理论的行列中。

粒子物理学当前的危机源于这样的事实：标准模型以外的理论分化为两个阵营，有的被证伪了，因而是错误的。其余的还未经检验——要么因为它们没有明确的预言，要么因为它们的预言还不能用现有的技术来检验。

在过去的30年，理论家们至少提出了十多个方法。每个方法都从一个令人信服的假定出发，但至今还没有一个成功的。在粒子物理学领域，这些方法包括“拟色”（technicolor）、“前子模型”和超对称

性。在时空物理学领域，有扭量理论、因果集、超引力、动力三角化[2]和圈量子引力。有些思想和它的名字一样怪异。

这些理论中，有一个吸引了绝大多数人的关注：弦理论。它流行的原因不难理解。例如，它声称正确描述了宏观和微观——大如引力，小如基本粒子；它还提出了所有理论中最大胆的假设：世界包含着看不见的维度和比我们知道的多得多的粒子。同时，它认为所有基本粒子源于同一个服从美妙的简单定律的实体——弦——的振动。它宣称统一了大自然所有的粒子和力。同样，它还许诺能澄清曾经做过或可能做的任何实验的预言。最近20年，人们为弦理论花费了很大的气力，但我们仍然不知道它是否正确。即使事情都做好了，理论也没有一个新预言能用今天的实验（哪怕今天所能想象的实验）来检验。它确实做出的几个明确的预言也是其他理论已经预言过的。

弦理论没有新预言，部分原因在于它似乎有数不清的形式。即使我们只考虑那些满足宇宙的几个基本事实（如宇宙的大小和暗能量的存在）的理论，我们要面对的弦理论还有10500（1后面跟500个零）个，比已知宇宙的原子还多呢。既然有那么多的理论，总有某个实验结果会满足其中的一个吧。因此，不论实验如何，弦理论都不可能被否定。但反过来也一样：没有实验能证明它是正确的。

同时，我们对多数弦理论都知之甚少；而我们多少知道些细节的少数几个理论，每一个通常都至少在两个方面不符合现有的实验数据。

于是我们面对着一个怪圈。我们知道如何研究的弦理论都已知是错误的，而我们无法研究的那些理论却有那么庞大的数目，真想象不出能有什么实验和它们全都不符。

问题还不仅于此。弦理论依赖于几个关键的假设，那些假设虽然有一定证据，但并没有证明。更糟糕的是，在经过了那么多艰辛的科

学劳动之后，我们仍然不知道是否存在一个完整而和谐的名叫“弦理论”的理论。其实，我们有的根本不是一个理论，而是一些近似计算的集合，外加一个猜想的网络——如果那些猜想正确，则意味着存在某个理论。但那理论还从没写出来过。我们不知道它的基本原理是什么，我们不知道该用什么数学语言来表述它——也许需要创造一门新的语言来描写它。既没有基本原理，也没有数学形式，我们又凭什么说我们知道弦理论宣扬了什么呢？

下面是弦理论家格林（BrianGreene）在他的《宇宙经纬》里说的：“即使今天，在它出现30年之后，多数弦的实践者仍然相信我们不能满意地回答一个最基本的问题：弦理论是什么？……多数研究者觉得我们现有的弦理论形式还缺乏我们在其他重大理论进步中看到的那种核心原理。”[3]因基本粒子物理学的成就而获诺贝尔奖的特胡夫特（Gerardt　Hooft）曾这样形容弦理论的现状：“实际上，我还没打算称弦理论是一个‘理论’，它更像一个‘模型’，甚至连模型也算不上，而只是一种感觉。毕竟，一个理论应该有一套指南，教我们如何识别我们想要描述的事物（在我们的情形，即基本粒子），而且至少在原则上能确立一些法则来计算那些粒子的性质，并做出新的预言。假定我给你一把椅子，却告诉你那椅子的腿还没找到，坐垫和靠背也要等会儿才拿来。那么我到底给了你什么？还能称它是椅子吗？”[4]因标准模型获诺贝尔桂冠的格罗斯（DavidGross）后来成为弦理论最激进、最令人敬畏的拥护者。不过，在最近一个庆祝弦理论进展的会议上，他这样结束自己的讲话：“我们不知道我们在谈论什么……物理学今天的状态就像我们为放射性感到疑惑的那个时候……它们还缺少某个绝对基本的东西。我们也许缺乏同样深刻的东西。”[5]

可是，尽管弦理论如此残缺，连其存在都是一个未经证明的猜想，做弦理论的人依然相信它是理论物理学前进的唯一道路。很久以前，有人请圣巴巴拉加州大学卡维里（Kavli）理论物理研究所的杰出的弦理论家波尔金斯基（JosephPolchinski）谈谈“替代弦理论”的问

题。他说，他的第一反应是，“这太无聊了，它没有替代者……所有好思想都是弦理论的一部分。”[6]哈佛助理教授莫特（LubosMolt）最近在他的博客上宣称，“没人能让别人相信弦理论的什么替代者，最可能的理由就是，也许弦理论不存在替代者。”[7]

怎么回事呢？在科学中，理论一词通常意味着非常确定的事物。兰多尔（LisaRandall）是莫特在哈佛的同事，一个有影响的粒子物理学家，她将理论定义为“嵌在一组关于世界的基本假设里的确定的物理学框架——一个囊括了众多现象的精简的框架。理论产生一组方程和预言，通过与实验数据的一致而得到证实。”[8]

弦理论不是这样的——至少现在还谈不上。那么，某些专家在连弦理论到底是什么都不知道的时候，又凭什么相信它不可替代呢？他们相信它不可替代究竟是什么意思？就是这些问题激发我写这本书。

理论物理学家难做，非常难做。不是因为需要很多的数学，而是因为存在很大的风险。当我们回顾当代物理学史的时候，会一次又一次地看到，做这样的科学不可能没有风险。如果很多人为一个问题奋斗了多年还没找到答案，那可能是答案太难而不那么显而易见，或者就是那问题没有答案。

就我们理解的说，弦理论认为世界与我们知道的根本不同。假如弦理论正确，世界该有更多的维、更多的粒子和更多的力。许多弦理论家在谈话或写作时，似乎都把额外的维和粒子的存在看成了确定的事实——任何优秀科学家都不能怀疑的事实。有个弦理论家不止一次地对我说，“你的意思是你认为可能没有额外的维？”事实上，不论理论还是实验，都没有任何证据表明存在额外的维。本书的目的之一就是要剥去弦理论的这层神秘的外衣。它的思想很美妙，出发点也好，但为了认识它为什么没有带来大的进步，我们必须清楚有什么支持的证据，有什么缺失的东西。

因为弦理论是高风险的事业——尽管学术界和科学社会支持，实验却并不支持——它的归宿只有两个。假如弦理论是正确的，理论家们会成为科学史上最伟大的英雄。凭着手头的一些线索——没有一条是确凿无疑的——他们将发现实在远比从前想象的广大。哥伦布发现了西班牙国王和王后不知道的新大陆（当然新大陆也不知道有西班牙皇家），伽利略发现了新恒星和卫星，后来的天文学家发现了新行星。所有这些发现在新空间维度的发现面前都将黯然失色。而且，许多弦理论家相信，众多弦理论所描述的无数世界确实存在着——那是我们不可能直接看见的宇宙。如果真是这样，那么我们所认识的实在，还不如任何洞穴人群所认识的地球。人类历史上还不曾有谁正确猜想过已知的世界会是那么广大。

另一方面，假如弦理论家错了，就不可能是小错。假如不存在新的维度和对称性，我们会将弦理论家归入失败者之流，就像开普勒和伽利略迈步向前时的那些还在研究托勒密本轮的人。他们的故事将警示人们什么不能算作科学，如何才能不让理论猜想超越了理性的极限而走进幻想。

因为弦理论的兴起，从事基础物理学研究的人们分裂为两个阵营。许多科学家继续做弦理论，每年大约有50个新博士从这个领域走出来。但还有些物理学家对弦理论深表怀疑——他们有的从来不看好它，有的则感到绝望，不再相信那个理论能有一个和谐的形式或做出什么真正的实验预言。分裂并不总是友好的。他们都怀疑对方的专业能力和道德水平，现在真正需要做的是维护两家的友谊。

根据我们在学校学过的科学图景，这样的情形不会持续下去。老师教我们，现代科学的总特征就是存在一个引导我们进一步认识自然的方法。偏差与争议当然是科学进步必须经历的，但人们总认为存在一种方法，通过实验或数学来解决争论。可是，在弦理论的情形，这样的机制似乎破产了。弦理论的许多信奉者和评论家们太相信他们的观点了，即使在朋友之间也很难展开诚恳的讨论。他们说，“你怎么

会看不见这个理论的美妙呢？一个如此美妙的理论怎么会不正确呢？”这也激起怀疑者们同样剧烈的反应：“你失去理智了吗？你怎么那么坚信随便的一个毫无实验证据的理论呢？你忘了怎么做科学吗？你连那理论是什么都不知道，怎么还如此确定你是对的呢？”

我写这本书就是希望能在专家和爱好者之间展开真诚而有益的讨论。不论我在这些年看到了什么，我还是相信科学。我相信科学家群体有能力摆脱挖苦和讽刺，通过基于当前证据的理性的讨论来解决争议。我很清楚，仅凭提出这个问题，我就会惹恼我的一些做弦理论的朋友和同事。我只能说，我写此书不是为了攻击弦理论或相信弦理论的人，我对他们满怀着敬意，我写它，首先是表达我对物理学的科学家群体的信任。

所以，这不是一本关于“我们”和“他们”的书。在我的经历中，既做过弦理论，也走过其他量子引力（爱因斯坦广义相对论与量子理论的融合）的道路。尽管我的主要努力在其他方法上，我也有一段时间很相信弦理论，而且试着去解决它的关键问题。虽然问题没解决，我还是写了18篇文章；所以，我要讨论的错误不仅是别人的，同样也是我自己的。我将谈到那些为大家广泛接受的猜想，尽管它们从没得到证明。我也曾相信过它们，因为相信才选择了我的研究方向。我还要说年轻科学家为了美好前程而追求那些公认的主流问题所受的压力。我本人深有体会，时常任由它们决定我的职业生涯。一方面需要独立的科学判断，一方面不能偏离科学主流，这样的冲突我也经历过。我写这本书不是想批评有不同选择的科学家，而是想明白为什么科学家需要面对那些选择。

实际上，我犹豫了很长时间才决定写这本书。我个人不喜欢冲突和对抗。毕竟，在我们从事的这门科学里，任何值得做的事情都是一种冒险，真正重要的是50年后我们的学生的学生认为哪些东西值得教给他们的学生。我一直希望身处弦理论研究中心的人能写一本客观而翔实的评论，告诉我们理论做到了什么，没做什么。没有人出来写。

之所以想把这些问题公开出来，是因为几年前在科学家与一群人文学者和社会科学教授（“社会建构主义者”）之间发生了一场争论，争论科学是如何进行的。社会建构主义者们声称科学团体并不比其他人类群体更理性和客观。多数科学家不是这样看科学的。我们教导学生，对科学的信仰必须基于对证据的客观估价。我们的对手反驳说，我们关于科学运行的主张不过是一种宣传，为的是向人们索求权利，整个科学事业和其他人类领域没有什么两样，都是在相同的政治和社会力量驱动下运行的。

我们科学家在论战中用的主要论据之一是我们的群体和他们不同，因为我们用很高的标准约束自己——根据那样的标准，任何一个理论如果未经公开的计算和数据的证明，未能消除专业人士的怀疑，我们是不可能接受它的。当然，正如我下面要更详细讲的，在弦理论中就不是这样的。尽管没有实验的支持和精确的形式，仍然有人相信它，似乎感情超过了理智。

弦理论的大肆宣扬使它成了探索物理学大问题的一条基本路线。在崇高的普林斯顿高等研究院享有永久职位的每个粒子物理学家几乎都是弦理论家，唯一的例外是几十年前来这儿的一位。在卡维里理论物理研究所也是如此。自1981年麦克阿瑟（Mac Arthur）学者计划开始以来，9个学者有8个成了弦理论家。在顶尖的大学物理系（伯克利、加州理工、哈佛、麻省理工、普林斯顿和斯坦福），1981年后获博士学位的22个粒子物理学终身教授中，有20个享有弦理论或相关方法的声誉。

弦理论如今在学术机构里独领风骚，年轻的理论物理学家如果不走进这个领域，几乎就等于自断前程。即使在宇宙学和粒子现象学等弦理论没有任何预言的领域，研究者们也常在讲话或文章的开头声称他们相信他们的工作将来可以通过弦理论推导出来。

有很好的理由把弦理论认真看作一种关于自然的假说，但这不等

于说它是对的。我为弦理论工作过几年，因为我那时非常相信它，很想凭自己的手解决它的关键问题。我也相信除非我像专业人士那样熟悉它，否则我没有资格发表意见。同时，我也在其他方向做过一些工作，它们也有希望回答基本问题。结果，有人怀疑我在争论时有点儿两面派。有些弦理论家说我“反弦”，这不能说一点儿不对。如果不是曾经为它着迷，感觉它可能成为真理的一部分，我是不可能花那么多时间和精力去做弦理论的，也就不会在它的问题的激发下写三本书。我不会支持科学以外的任何事情，也不会反对任何事情，除非它威胁了科学。

但是，除了和同事的关系而外，还有很多事情也出了问题。为了研究，我们物理学家需要重要的资源，那主要来自我们的公民——通过税收或者基金。他们需要的回报只是希望有机会通过我们的肩膀看看我们是如何前进的，深化了多少关于我们共同的世界的知识。与公众沟通的物理学家，不论通过写作、谈话、电视还是网络，都有责任说出真相。我们必须谨慎地在讲成功的同时也讲失败。其实，真诚面对失败不会损害我们的事业，反而会促进它。毕竟，支持我们的人生活在现实世界，他们理解任何事业的进步都需要真正的冒险，也都可能遭受失败。

近年来，许多面向大众的图书和杂志都描述过理论物理学家正在探讨的新思想。有些叙述根本就没有用心解释新思想距离实验验证和数学证明还差多远。公众渴望知道宇宙的奥秘，我觉得有责任保证本书讲的故事都尊重事实。我希望把我们不能解决的各种问题都呈现出来，说清楚实验支持什么，不支持什么，并把事实与猜想和智力的时尚区别开来。

首先，我们物理学家要对我们的未来负责。正如我后面要说的，科学基于一定的道德规范，而道德规范要求其实践者必须诚实。它还要求每个科学家应该是他所信仰的事物的法官，于是每个未经证实的思想在被证明之前将面对众多怀疑和批评。这反过来要求我们应该欢

迎不同的方法走进科学群体。我们做研究是因为即使我们中最聪明的人也不知道问题的答案。答案经常在主流以外的某个方向。在那些情形，即使主流猜对了，科学的进步也需要那些抱不同观点的科学家们的支持。

科学需要在同一与多样之间达成微妙的平衡。因为我们很容易欺骗自己，因为答案未知，不论多么训练有素、多么精明的专家，都可能反对即将成功的方法。因此，为了科学的进步，科学群体必须支持任何一个问题的不同方法。

很多证据表明，这些基本原则在基础物理学中不再受人尊重了。虽然大家都赞成多样性的说法，做起来却不是那样。有些年轻的弦理论家告诉我，不论相信与否，他们感觉做弦理论很压抑，因为它在大学里成了通向教授职位的入场券。他们说的是真的：在美国，追求弦理论以外的基础物理学方法的理论家，几乎没有出路。最近15年，美国的研究型大学为做量子引力而非弦理论的年轻人一共给了三个助理教授的职位，而且给了同一个研究小组。当弦理论还在为科学争斗时，它在学术圈里已经赢了。

这伤害了科学，因为它抑制了其他方向的考察，而有些方法是很有希望的。尽管这些方法的投入不够多，但有的已经赶在了弦理论的前头，可能为实验提出了明确的预言，目前正在进行中。

1000多个最聪明、受过最好教育的科学家在最好的条件下研究的弦理论，怎么会面临失败的危险呢？这令我困惑了多年，不过现在我想我知道答案了。我认为将要失败的不是什么特别的理论，而是一种科学作风，它适合我们在20世纪中期遇到的问题，却不适合我们今天面临的基本问题。粒子物理学的标准模型是一种特殊科学作风的成功，从而在40年代独领风骚。它迥然不同于爱因斯坦、玻尔、海森伯、薛定谔和20世纪初的其他科学革命者们的科学作风。他们的工作源于时间、空间和物质的最基本问题，他们将自己的工作视为他们所

熟悉的更广大的哲学传统的一部分。

在由费曼、戴森等人发展起来的粒子物理学方法户，不需要再考虑基本的问题，这就使他们从量子物理学意义的争论中解放出来，而他们的前辈却一直纠缠于那些争论，并带来了30年的飞速进步。事情本应如此：解决不同的问题需要不同的研究风格。发挥已有框架的应用，与起初搭建那些框架相比，需要截然不同的思想和思想者。

不过，正如我下面要讲的，过去30年的教训是，我们眼下的问题不可能通过那种实用主义的科学路线来解决。为了让科学不断进步，我们要再次面对空间、时间和量子理论的基本问题。我们会看到，科学进步的方向是将理论带回实验来检验，引领这个方向的人，赶上了发现新思想比追赶新潮流更容易的时代，而且他们主要也是以20世纪初的先驱者们的风格做科学。

我想强调的是，我并不针对弦理论家个人，我知道他们有些是卓有成就的物理学家。我会第一个站出来说，他们有权利追求他们认为最有希望的研究。但我非常担心那样一种趋势，大家都在支持某一个研究方向，却冷落了其他有希望的方向。

正如我想说的，假如真理所在的方向需要我们重新思考关于空间、时间和量子世界的基本思想，那么那种趋势可能带来悲剧性的后果。

# 致谢

每本书都先有一个想法，而本书的萌芽要感谢John Brock-man，他发现我想就民主与科学问题做点儿事情，而不是写一本模糊的学术专著。那是本书的一个主题，但正如他预见的，在特殊的科学争论的背景下，论证会显得更加有力。我要感谢他和Katinka Mastson，他们一直支持着我，请我加入第三文化组成的群体。他们为我提供了我的专业之外的氛围，改变了我的生活。

没有哪个作者能遇到比Amanda Cook更好的编辑了，如果说本书有什么好的地方，那要归功于她的指导和参与。SaraLippin-cott以任何作者都羡慕的优美和准确做好了一切。和她们合作是我的荣幸。Holly Bemiss, WillVincent和HoughtonMifflin的每个人都满怀热情和技巧地关注着这本书。

在过去几十年，许多同事都花时间教我弦理论、超对称和宇宙学。其中，我要特别感谢Nima Arkani-Hamed, Tom Banks, Mi-chaelDine, JacquesDistler, MichaelGreen, BrianGreene, Gary Horowitz, CliffordJohnson, RenataKallosh, JuanMaldacena, Lubos Motl, HermannNicolai, AmandaPeet, MichaelPeskin, JoePolchins-ki, LisaRandall, JohnSchwarz, SteveShenker, PaulSteinhardt, Kel-logg Stelle, Andrew Strominger, Leonard Susskind, CumrunVafa和Edward Witten，感谢他们的时间和耐心。如果说我们还有不同的意见，我希望本书不是最后的陈述，而只是小心组织的论证，是我为正在进行的对话的一点贡献——我是怀着对他们的努力的敬意和尊重参与进来的。如果证明世界是11维超对称的，我会第一个站出来欢呼他

们的胜利。但是现在我要预先感谢他们允许我解释，为什么我经过长期思考以后不再相信有那种可能。

这不是一段学术史，但我确实在讲历史，几个朋友和同事慷慨地花了很多时间给我讲了真实的故事而不是久远的传说。Julian Barbour, JoyChristian, Harry Collins, JohnStachel和AndreiStari-nets给我整个手稿加了专业注释。为了使本书尽可能浅显易读，我对注释进行了选择，产生的错误当然是我个人的责任。错误的修正和进一步的思考将贴在本书的网页上。读过手稿并提出过批评的其他朋友和家庭包括Cliff Burgess, Howard Burton, Margaret Geller, JaumeGomis, Dina Graser, Stuart Kauffman, JaronLanier, Janna Levin, Joao Magueijo, Patricia Marino, FotiniMarkopoulou, Carlo Rovelli, Michael Smolin, PaulineSmolin, RobertoMangabeira Unger, Antony Valentini和Eric Weinstein.ChrisHull, JoePolchins-ki, Pierre Ramond, Jorge Russo, MosheRozali, John Schwarz, An-drew Strominger；Arkady Tseytlin也帮我澄清了一些具体的事实和问题。

多年来，我的研究得到了国家科学基金会的资助，对此我非常感谢。但我无比幸运地遇见了一个人，他问我，“你真的想做什么？你最疯狂、最雄心勃勃的思想是什么？”接着，出乎我的意料，Jeffrey Epstein给了我寻找我自己的答案的机会，令我难忘。

本书的部分内容是谈有关科学共同体的价值，我很幸运地从量子时空的先驱者那里知道了我的价值：Stanley Deser, David Finklestein, JamesHartle, ChrisIsham和RogerPenrose。如果没有别人的支持和合作，我不会有任何结果。特别是AbhayAshtekar, JulianBarbour, LouisCrane和CarloRovelli。我还要感谢我最近的合作者Stephon Alexander, MohammadAnsari, OlafDreyer, Jerzy Kowalski-Glikman, JoaoMagueijo，特别是FotiniMarkopoulou，她不断的批评和挑战令我保持冷静而不自以为是。还应该说明的是，如果没有更多的不赶潮流而投身物理学基本问题的物理学家、数学家和

哲学家，我们的工作是不会有意义的。本书首先要献给他们。

如果没有朋友们的支持，我的工作和生活将贫困潦倒；他们不但让我做科学，也让我在更广大的背景下认识了科学。这些人包括SaintClairCemin, JaronLanier, DonnaMoylan, ElizabethTurk和MelanieWalker。

每本书都是带着地域的精神写作的。我最早的两本书是纽约和伦敦。这本书则带着多伦多的精神。Pico Iyer称它为未来之城，我想我幸运地知道那是为什么。在2001年9月的某个时候，我一个外乡人来到这里，受到了大家的欢迎。我首先要感谢Dina Gra-ser，当然还有CharlieTracyMacdougal, OliviaMizzi, Hanna Sanchez和港湾帆船俱乐部（如果你去年春天没在水上看见我，就是因为它！）的其他伙伴。

我要感谢Howard Burton和Mike Lazaridis请我到这里来。他们建立了PI理论物理研究所，对科学的远见和支持，我不知道还有什么更伟大的行动。他们真诚地相信科学的未来，一直投身于研究所的成功，关心科学的人应该对他们报以最高的赞美。我特别感谢他们为我提供了一个机会，那是我个人的机会，也是科学的机会。

我们同甘共苦建立了研究所和研究团队，我要感谢Clifford Burgess, FreddyCachazo, LaurentFreidel, JaumeGomis, DanielGot-tesman, LucienHardy, JustinKhoury, RaymondLaoamme, Fotini Markopoulou, MicheleMosca, RobMyers, ThomasThiemann, Antony Valentini以及其他敢于冒险将自己的学术生涯投入到我们的艰苦历程中来的人，可惜不能列举他们的名字。尽管可以不说，我还是想强调一句，本书的每句话都是我自己的观点，绝不代表PI及其科学家和创建者们的正式或非正式观点。相反，本书成为可能，得益于我身在一群特别的科学家中间，他们提倡真诚的科学争论，他们知道，在科学的进程中，活跃的讨论不需要像朋友那样相互扶持。如果有更多像PI

这样的地方，我就不会觉得有必要写这本书了。

最后，我感谢父母给我的无私的爱和支持，感谢Dina，把一切都做得井井有条，令我感到了生活的乐趣，也让本书讨论的东西井然有序。

# 第一篇　未完成的革命

## 第一章　理论物理学的五大问题

从物理学萌芽以来，就有人想象自己是最后一代还会面对未知的人。物理学在它的实践者们看来几乎就要圆满了。当诚实的人被迫承认他们对基础还一无所知时，这种满足就在革命中破灭了。可是，就连那些革命也仍然幻想着就在某个角落藏着我们需要的大思想——它能整合我们对知识的追求，并将圆满地终结它。

我们就生活在那样的一个革命时期，而且经过了一个世纪。最近的一个时期是哥白尼革命，从16世纪初兴起，它颠覆了亚里士多德关于时间、空间、运动和宇宙的理论。革命的高潮是牛顿1687年出版的《自然哲学的数学原理》，他提出了物理学新理论。当前的物理学革命从1900年开始，普朗克在那年发现了描述热辐射谱能量分布的公式，证明能量不是连续的而是量子化的。这场革命还没有结束。物理学家今天需要解决的问题还没有答案，在很大程度上就是因为20世纪的革命不够彻底。

我们没能完成这场科学革命的主要原因在于五个问题，每一个都很棘手。我从20世纪70年代开始做物理时就面对这些问题了；在过去的30多年里我们对它们有了很多的认识，但还是不能解决。任何基础物理学理论，不论什么样子的，都必须解决这五个问题，所以我们需要好好看看它们。

爱因斯坦当然是20世纪最重要的物理学家。他最伟大的成就也许是发现了广义相对论，也是迄今为止我们最好的关于时间、空间、运

动和引力的理论。他深刻地洞察到引力和运动是密切联系的，而且联系着空间和时间。这个思想打破了几百年的时空概念，那时人们一直以为时间和空间是固定而绝对的。正因为时空永恒不变，所以它们成了我们过去用于定义运动概念（如位置和能量）的背景。

在爱因斯坦的广义相对论中，空间和时间不再充当固定不变的背景。空间与物质一样是动态的，既有运动，也有变形。结果，整个宇宙既可能膨胀，也可能收缩，而时间甚至可能有开始（在大爆炸）和结束（在黑洞）。

爱因斯坦还有其他的贡献。他是第一个认识到需要新的物质和辐射理论的人。诚然，普朗克公式隐含着需要突破，但普朗克对其意义的认识还不够深刻，他觉得那可以与牛顿物理学协调起来。爱因斯坦的想法相反，他在1905年第一次明确地论证了那样一个理论。20多年后，那个理论才出现，就是我们熟悉的量子理论。

这两个发现（相对论和量子论）都需要我们与牛顿物理学彻底决裂。然而，两个理论尽管在过去100年里取得了伟大的进步，却依然不够圆满。每个理论都有缺陷，意味着存在一个更深层的理论。不过，我们说每个理论都不完备，主要还是因为两者的并存。

因为简单的理由，我们的思想呼唤着第三个理论来统一所有的物理学。大自然显而易见是“统一的”。我们所在的宇宙是相互联系的，因为万物都发生相互作用。我们决不能有两个分别覆盖不同现象而毫不相干的理论。任何所谓的终极理论都必须是一个完备的自然理论，应该囊括所有我们知道的东西。

没有那样的统一理论，物理学也延续了那么长久。原因是，就实验而言，我们可以将世界划分为两个领域。在量子物理统治的原子领域，我们通常可以忽略引力的作用，像牛顿那样将空间和时间看做不变的背景。另一个领域则属于引力和宇宙学，我们可以在那个领域里

忽略量子现象。

可是这顶多不过是一种临时的权宜之计。要超越它，是理论物理学中的第一个大问题。

问题1：将广义相对论与量子理论结合为一个真正完备的自然理论。

这就是所谓的量子引力问题。

除了基于自然统一性的理由，两个理论还有各自的具体问题，也需要彼此的统一。每个理论都有无穷大问题。在自然界，我们还没有遭遇过具有无穷大数值的东西。但我们在量子理论和广义相对论中都预言过有物理意义的量变成了无穷大。这也许是大自然在惩罚那些胆敢破坏她统一的理论家。

广义相对论的无穷问题在于黑洞内部的物质密度和引力场强度会很快变成无穷大。在宇宙历史的极早期可能就是这样的——假如我们真的相信广义相对论正确描述了宇宙的鸿蒙时代。在密度成为无穷大的点，广义相对论方程破灭了。有人解释这是时间停止了，但更冷静的观点认为那是因为理论不够完备。长期以来，明智的人都猜想不完备的原因在于忽略了量子物理学的效应。

反过来，量子理论也有自己的无穷大困惑。每当我们用量子力学来描述场（如电磁场），它们就会出现。这儿的问题在于电场和磁场在空间的每一点都有确定的数值，这意味着有无限多个变量（即使在有限体积的空间里也有无穷多个点，因而有无穷多个变量）。在量子理论中，每个量子变量的值都存在无法控制的涨落。无穷多个变量加上无法控制的涨落，就可能带来那样的方程——当我们寻求某个事件发生的概率或某个力的强度时，它们会摆脱我们的掌握，产生无穷大的结果。

所以，在这种情形中，我们仍然不得不感觉物理学失去了一个基本的部分。很久以来，人们都希望把引力考虑进来就可能平息涨落，全都成为有限。如果说无穷大是缺失统一的标志，那么统一的理论就不会出现任何无穷大。我们称那样的理论为有限理论，即它用合理的有限的数回答了所有的问题。

量子力学成功解释了大量现象，其领域包括从辐射到晶体管性质，从基本粒子物理学到作为生命基本组成的酶和其他大分子的行为。在过去的100年里，它的预言经历了一次又一次的证明。但有些物理学家对它总怀疑虑，因为它描述的实在太奇异了。量子理论内部包含着明显的概念性疑难，在它诞生80多年以后仍然没得到解决。电子既像粒子也像波，光也如此。而且，理论对亚原子粒子只能给出统计的预言。因为不确定性原理的限制，我们也不可能做得更好。那原理说，我们不能同时测量粒子的位置和速度。理论只能得到概率。一个粒子（如电子）在我们观测之前可以处于任何位置；从某种意义说，是我们的观测决定了它的状态。所有这些都说明量子理论没有告诉我们完整的故事。结果，不论它多么成功，还是有很多专家相信量子理论隐藏了自然的某些基本的东西，而那是我们需要知道的。

从一开始就困扰量子理论的是关于实在与形式的关系问题。物理学家一贯希望科学应该说明我们之外的实在。物理学不仅是一堆预言我们在实验中看到什么的公式，而且应该为我们提供实在本来的图像。我们是远古灵长类动物的偶然的后代，最近才出现在悠远的世界历史长河中。实在不可能依赖于我们的存在。没有观众的世界的问题也不可能请外星文明来回答，因为世界曾经致密而火热，不可能形成智慧生命。

哲学家称这种观点为“实在论”，可以概括为一句话：实在的世界必然独立于我们而存在。因此，科学描述实在的方法不能以任何基本的方式涉及我们选择测量什么，不测量什么。

量子力学——至少第一次出现的那种形式——并不容易满足实在论。这是因为理论预先把自然分为两个部分。界线的一边是有待观测的系统，而我们观测者在另一边。我们拥有的是做实验的仪器、测量的工具，还有确定事件发生的时钟。量子理论可以说是一门新的语言，它沟通我们和我们用仪器研究的系统。这门量子语言的动词是我们的实验和观测，名词是观测到的东西。它并不告诉我们没有我们的世界是什么样子的。

自量子理论第一次出现以来，在接受与不接受它的科学作风的两派人物之间，一直存在着论战。量子力学的许多创立者，包括爱因斯坦、薛定谔和德布罗意发现这种物理学方法令人厌倦。他们是实在论者。在他们看来，量子理论不论表现多好，都是不完备的理论，因为它不能提出一幅没有我们的相互作用的实在图景。论战的另一方是玻尔、海森伯等许多人，他们没有为这种科学方法感到惊骇，而是热情地拥抱它。

从那时起，实在论者一直指责量子理论的现有形式存在着矛盾。有些矛盾是显然的，因为，如果量子理论是普适的，那它也该描述我们。于是问题来了：为了明白量子理论的意义需要把世界一分为二。困难在于把那条分界线画在什么地方。这依赖于谁在进行观测。当你测量一个原子时，你和你的仪器算一边，而原子算另一边。但是，假如我通过我设置在你的实验室里的摄像头观察你的工作，那么我可以把你的整个实验室——包括你和你的仪器以及你正在观测的原子——看作我要观测的系统。而另一边只有我。

于是，你我描述了两个“系统”。你的系统只有原子，而我的系统包括你、原子和你用以研究原子的所有东西。你所观测的东西在我看来是两个相互作用的物理系统。因此，即使你同意把观测者的行为作为理论的一部分，那理论也是不充分的。量子力学需要扩展，需要容纳很多不同的、依赖于观测者的描述。

整个问题归结为量子力学的基础问题。这是当代物理学的第二个大问题。

问题2：解决量子力学的基础问题：要么弄清理论所代表的意义，要么创立一个新的有意义的理论。

解决这个问题有几条不同的路线。

1.为理论找一种语言，它能解开上面提到的所有疑难，并将系统与观测者的世界划分作为理论的一个基本特征。2.以实在论的观点重新解释理论——重新解读方程，从而使测量和观测在描述基本实在中不再起作用。

3.创立新理论，提出比量子力学更深刻的对自然的认识。

这三条途径，眼下都有一群聪明的人物在追寻。遗憾的是，没有多少物理学家研究这个问题。有时，这说明问题已经解决或无关紧要。可现在不是这样。它可能是现代科学所面临的最严峻的问题，只是因为问题太难，所以进步缓慢。我敬佩那些研究这个问题的人，他们怀着纯洁的目标，勇敢地蔑视潮流，向着最困难也最基本的问题进攻。

可是，他们卓越的工作也没能解决问题。在我看来，这意味着那不仅仅是找一条新路来思考量子理论的问题。创立理论的人不是实在论者，他们不相信人类能构造一幅真正的独立于我们的行为和观察的世界图景。相反，他们赞成不同的科学观：科学只不过是我们用以描述我们的行为和观察的寻常语言的延伸。

后来，那种观点似乎太放纵了——那也是时代的产物，因为我们曾希望我们已经在很多方面有了超越。那些继续捍卫量子力学，坚持认为它是一个世界理论的人，也高举起实在论的大旗。他们赞成沿着实在论的路线重新解释理论。然而，当他们提出有趣的建议时，却没

有一个完全令人信服。

实在论作为一种哲学也许会消亡，但不大可能。毕竟，实在论是科学家前进的动力。对我们多数人而言，相信那个独立于我们的实在，相信我们能真正理解它，这样的信念驱动着我们去做艰辛的工作，做一个科学家，去认识大自然。面对实在论者清理量子理论的失败，我们越发觉得唯一的选择可能是第三个：找一个更遵从实在论解释的新理论。

我承认我是实在论者。我站在爱因斯坦等人一边，相信量子力学是对实在的不完整描述。我们从哪儿去寻找量子力学缺失的东西呢？在我看来那需要我们更深入地理解量子物理学本身。我相信，如果到头来还不能解决问题，那是因为还有确实的关联着其他物理学问题的东西。量子力学的问题不会孤立地解决；当我们统一物理学的努力取得重大进步时，它们也许会迎刃而解。

如果真是这样，它在两个方面都有用：只有当我们发现了量子力学的合理替代者时，才可能解决其他的大问题。

物理学应该统一，这个思想比任何其他问题对物理学研究的驱动更大。但物理学可以通过不同的方式来统一，我们要谨慎区分。迄今为止我们都在讨论用一个定律来统一。很难想象有谁能否定这是必须的目标。

但还有其他方式来统一世界。爱因斯坦——当然，他和任何人一样，也是这样想的——强调我们必须区分两类理论：原理理论和构造性理论。原理理论搭建自然描述的框架。根据定义，原理理论一定是普适的，能用于天下万物，因为它确立了我们用于描述自然的语言。不可能有分别用于不同领域的两个不同的原理理论。因为世界是统一的，任何事物最终都与其他事物相互作用，因此只能用一种语言来描述那些相互作用。量子理论和广义相对论都是原理理论。所以，逻辑

需要它们的统一。

另一类理论，即构造性理论，以明确的模型或方程描述特殊的现象。[9]电磁场理论和电子理论是构造性理论。这种理论不可能是孤立的，必然建立在原理理论的框架下。但只要原理理论允许，可以存在服从不同定律的现象。例如，电磁场服从的定律就不同于假想的宇宙暗物质（远远超过宇宙中寻常的原子物质的数量）的定律。关于暗物质，我们知道一点：不管它是什么，它都是“暗的”。这意味着它不发光，因此可能不与电磁场发生作用。这样，两个不同的理论可以和谐共存。

问题在于，电磁学定律不管世界还存在其他什么。不论夸克、中微子还是暗物质，可以有，也可以没有。同样，描述强弱两种在原子核中发生作用的力的定律也不必要求电磁力的存在。我们很容易想象一个只有电磁力而没有强力（或者相反）的世界。就我们现在的认识，那样的世界也是和谐的。

但我们仍然可以问，是否我们所看到的所有自然力都表现为单独的基本力。我只能说，似乎没有逻辑论证应该如此，不过那可能是对的。

统一各种力的愿望，在物理学的历史上带来了几个重要的进步。麦克斯韦在1867年将电和磁统一为一个理论，一个世纪后，物理学家发现电磁场与传播弱核力（即引起放射性衰变的力）的场也能统一。这就引出弱电理论，其预言在过去30年得到了一次又一次的证明。

在电磁场与弱场的统一之外，自然还有两种基本力（我们知道的）。那就是引力与强核力（将夸克结合在一起形成质子和中子的力）。这四种力能统一起来吗？

这就是我们的第三个问题：

问题3：确定不同的粒子和力能否统在一个理论并将其解释为一个单独的基本作用。

为了区别上面讨论的定律的统一，我们不妨称这个问题为粒子和力的统一。

乍看起来，这个问题很容易。1918年人们就提出了第一个统一引力和电磁力的建议，以后越来越多。只要我们忘记自然是量子力学的，这些理论表现都不错。如果把粒子物理学从图景里拿出去，统一理论是很容易构造的。但如果要把量子理论包括进来，问题就困难得多。因为引力是四种基本自然力之一，我们必须在解决这个统一问题的同时解决量子引力的问题（即问题1，融合广义相对论与量子理论）。

在上个世纪，我们略微简化了对世界的物理描述。就基本粒子而言，似乎只有两类：夸克和轻子。夸克是质子和中子的成分，我们发现的许多粒子也和它们相似。轻子一组囊括了所有非夸克构成的粒子，包括电子和中微子。总之，已知的世界可以用六种夸克和六种轻子来解释，它们通过四种力发生相互作用：引力、电磁力、强核力和弱核力。

12种粒子和4种力就是我们解释世界万物所需要的一切。我们对这些粒子和力的基本物理学也十分清楚。我们的认识形成的那个理论，解释了所有的粒子和所有的力，但引力除外。那就是基本粒子的标准模型——简称标准模型。这个理论没有先前说的无穷大疑难。我们想计算的任何东西，都能得出有限的结果。自理论建立30多年来，许多预言都经历了实验的考验，每次都得到了证实。

标准模型建立于20世纪70年代初。除了发现中微子有质量而外，它不需要什么修正。那么，物理学为什么到1975年还没大功告成呢？还有什么没做呢？

标准模型尽管有用，却有个大问题：可调节的常数太多。当我们陈述理论的一个定律时，必须确定这些常数的值。可根据我们当前的认识，任何数值都可以，因为不管我们为常数赋以什么数值，理论在数学上都是和谐的。这些常数决定了粒子的性质。有的决定夸克和轻子的质量，有的决定力的强度。我们不知道为什么常数有那些值；我们只是做实验把它们确定下来。如果你把标准模型看做一台计算机，那么常数就是键盘，你可以把它放在你喜欢的任何地方，而不会影响程序的运行。

这样的常数大约有20个，一个所谓的基本理论有那么多可以自由调节的常数，是很令人尴尬的事情。每个常数都代表某个我们不知道的基本事实：什么物理原因或机制决定着那些常数有那样的观测值？

这是我们的第四个大问题：

问题4：自然是如何选择量子物理标准模型中的自由常数值的？

我们真诚地希望某个真正的粒子和力的统一理论能给问题一个唯一的答案。

1900年，英国大物理学家汤姆逊（William Thomson，即开尔文勋爵）曾公开宣称物理学到头了，例外的只是漂浮在远方地平线上的两小朵乌云。那“乌云”的背后竟然藏着量子论和相对论。现在，尽管我们欢呼标准模型加广义相对论囊括了所有已知现象，我们仍然明白还有两朵乌云，即暗能量和暗物质。

除了它与量子的关系问题，我们认为已经很了解引力了。从地球的自由落体和光，到行星及其卫星的运动，再到星系和星系团的尺度，所有这些观测都在很高的精度上证明了广义相对论的预言。过去认为奇异的现象——如引力透镜（物质对空间的弯曲效应）——现在我们已经习以为常，而且用来测量星系团的质量分布。

在许多情形——速度远小于光速和物质不太致密的情形——牛顿的引力定律和运动定律是广义相对论预言的极好近似。它们似乎也应该帮助我们预言，星系的物质和众多恒星如何影响某颗特别的恒星的运动。其实不能。牛顿的引力定律认为，任何环绕其他物体的物体的加速度正比于它所环绕的物体的质量。恒星质量越大，环绕它的行星的运动就越快。就是说，假如两颗恒星各有一颗行星环绕，而且行星距离恒星一样远，那么环绕大质量恒星的行星运动更快。因此，如果知道轨道上的天体的速度和它到恒星的距离，我们就能计算恒星的质量。同样的逻辑也适用于环绕星系中心运动的恒星。测量恒星的轨道速度，就能计算星系的质量分布。

在过去的几十年里，天文学家做过一个很简单的实验，以两种不同的方式测量星系的物质分布并比较其结果。首先，他们通过观测恒星的轨道速度决定星系的质量；其次，他们通过直接计数他们能看到的星系的所有恒星、气体和尘埃来确定其质量。他们的思路是比较两个结果。两个结果都应该说明星系的总质量和物质的分布。根据我们对引力的充分认识，加上所有形式的物质都发光，两种方法应该是一致的。

结果却不一致。天文学家比较了两种方法对100多个星系的测量结果。几乎所有情形下，两种测量都不一致，不止差一点儿，而是差了10倍。而且，差别总是指向同一个结论：为了解释观测到的恒星运动，直接计数的恒星、气体和尘埃是远远不够的，还需要更多的物质。

这只能有两种解释。也许第二种方法错了，因为星系的物质比可见的物质多得多；也许牛顿定律不能准确预言恒星在星系引力场中的运动。

我们所知的所有物质形式都会发光，要么像恒星那样直接发光，要么像行星、星际岩石、气体或尘埃那样反射光。所以，如果有我们

看不见的物质，它一定是某种新奇的物质形式，既不发光，也不反光。因为偏差巨大，所以星系物质的大多数必然是那种新形式的物质。

今天，多数天文学家和物理学家都相信这是对那两种测量差别的正确解释。丢失的物质原来是我们看不见的物质。这种神秘的丢失的物质就是我们说的暗物质。大多数人偏向暗物质假说，因为另一种唯一的可能是牛顿定律和广义相对论错了——那就太可怕了。

事情越发神秘了。我们最近发现，当我们对更大尺度（相应为数十亿光年）进行观测时，即使加入暗物质，广义相对论方程也不能满足。137亿年前的大爆炸所驱动的宇宙膨胀似乎正在加速，而根据我们看见的物质加暗物质，它应该是减速的。

仍然有两种可能的解释。也许广义相对论真的错了。它只在太阳系和我们银河系内部的邻近系统得到证明。如果扩大到整个宇宙的尺度，广义相对论可能不再适用了。

也许还有一种新的物质或能量形式（想想爱因斯坦著名的方程E＝mc2，能量与物质是等价的），将在那样的大尺度上发生作用。就是说，这种新能量形式只影响宇宙的膨胀。这样的话，它不可能聚积在星系甚至星系团的周围。我们为了观测数据而假设的这种奇异的新能量，叫暗能量。

多数物质都经受压力，而暗能量经受张力——就是说，它将物质拉拢而不是推开。因为这一点，张力有时也称负压力。虽然暗能量经受张力，它还是使宇宙膨胀更快。这一点令人困惑，是可以理解的。人们会认为负压力的气体像一根连接星系的橡皮圈，使膨胀速度慢下来。但结果证明负压力太强了，以致在广义相对论中出现了相反的效应。它引起宇宙的加速膨胀。

最近的观测表明宇宙主要是这种未知的物质构成的。足足有70%的物质密度似乎来自暗能量形式。26%是暗物质。只有4%是普通物质。因此，我们实验观测和粒子物理学标准模型描述的物质，还不足二十分之一。对其余的96%，除了刚才说的那些性质而外，我们一无所知。

最近10年，宇宙学观测更加精确了。这部分是因为摩尔律的作用——那个定律说，大约每18个月，电脑芯片的运行速度就提高一倍。所以新实验都用微芯片，要么用于卫星，要么用于地面的望远镜。所以芯片越好，观测也越好。今天我们对宇宙的基本特征（如总物质密度和膨胀速率）有了很多认识。我们还有一个像基本粒子物理学的标准模型一样的宇宙学标准模型。宇宙学标准模型也同样有很多自由调节的常数——大约15个。这些常数特别决定着不同类型的物质和能量的密度以及宇宙的膨胀速率。没人知道为什么常数会有那些值。和粒子物理学的情形一样，常数值是从观测得到的，还没有任何理论能解释。

这些宇宙学之谜构成我们的第五个大问题。

问题5：解释暗物质和暗能量。或者，假如它们不存在，那么该如何在大尺度上修正引力理论，为什么修正？更一般地说，为什么宇宙学标准模型的常数（包括暗能量）具有那样的数值？

以上的五个问题代表了我们当前认识的边界。它们令理论物理学家寝食难安，驱动着理论物理系的大部分前沿研究。

任何充当自然的基本理论的理论都必须回答这些问题的每一个。本书的目的之一只是估量最近的物理学理论（如弦理论）朝着那个方向走了多远。不过在那之前，我们还是先看看早期的物理学统一做了些什么。从那些成功（当然也有失败）中，我们能学会很多东西。

# 第二章　美的神话

物理学的夙愿和拙劣的爱情小说一样，是为了统一。在可能的时候，把过去认为不同的两样事物结合起来，作为一个实体的不同方面——这是科学最大的惊奇和快乐。

对假想的统一，人们最正常的反应就是惊讶。太阳不过是一颗恒星——恒星都是太阳，只是离我们太远而已！想象一下，16世纪末的铁匠或演员听到布鲁诺（Giordano Bruno）的这种奇谈怪论会有什么反应。还有比把太阳和恒星混为一谈更荒唐的事情吗？人们早就听说，太阳是上帝为了温暖地球而创造的一团伟大的烈火，而恒星不过是天球上的小孔，好让天国的光线透过。统一立刻就颠覆了我们的世界。我们曾经相信的成了不可能的。假如恒星是太阳，宇宙就会比我们想象的大得多！天国不可能就在我们的头顶！

更重要的是，新的统一构想还会带来从前无法想象的假说。假如恒星是别的太阳，必然有行星环绕着它们，那儿一定生活着其他的人类！这些话的含义往往会超出科学。假如存在生活着其他人类的行星，耶稣也许会去所有的星球，那么他降临人间就不是独一无二的事件；否则那些星球的人就不可能得到拯救！难怪天主教会要烧死布鲁诺。

大统一成了整个新科学赖以建立的思想基础。有时，结果会极大威胁我们的世界观，人们惊奇过后会立刻表示不信。在达尔文之前，每个物种都是永恒不变的，都是由上帝一个个创造出来的。但通过自然选择的进化，所有物种都有共同的祖先。它们统一在一个大家庭里。达尔文之前和之后的生物学简直不像是同一门学科。

如此强力的新观点很快引出新的发现。如果所有生命有共同的祖先，它们一定是以相似的方式创造的。的确，我们真是相同材料做的，因为所有生命都由细胞组成。植物、动物、真菌和细菌看似各不

相同，其实只不过是不同方式排列的细胞群。构造这些细胞、为它们提供能量的化学过程，在整个生命王国都是相同的。

如果统一的思想对从前的思维方式是一种巨大的震撼，为什么还有人相信呢？从多方面看，这正是我们要讲的问题，因为我们的故事讲的就是几个统一的思想，其中有的赢得了科学家们的强烈信心。但它们没有一个在科学家中达成共识。于是，我们常发生激烈的争论，有时甚至是斗气，这就是剧烈改变世界观的结果。那么，当人们提出某个统一思想时，我们又凭什么说它是对还是错呢？

可以想象，并非所有的统一思想都是正确的。有个时期，化学家提出热是一种物质，和任何普通物质一样。那种东西被称作热素。这个概念统一了热与物质，然而它是错的。热与物质的真正统一在于热是原子的随机运动的能量。尽管古代印度和希腊的哲学家已经提出过原子论，但直到19世纪后期，热作为原子随机运动的理论才真正发展起来。

在物理学历史上，出现过许多统一理论，后来证明都错了。一个著名的思想是，光与声音在本质上是一样的：它们都被认为是物质的振动。因为声音是空气中的振动，所以光也被设想为某种叫做以太的新物质形式的振动。正如我们的周围充满着空气，宇宙也充满着以太。爱因斯坦以他自己的统一方式推翻了这种奇异的思想。

理论家们过去30年研究的重要思想，如弦理论、超对称、高维时空、圈引力等等，都是关于统一的建议。我们凭什么说哪个对哪个错呢？

我已经讲过成功的统一理论具有的两个特征。第一是惊奇，不能低估了这一点。如果没有惊奇，那思想要么无聊乏味，要么就是我们以前知道的。第二是戏剧性的结果：统一应该立刻引出新的发现和假说，从而成为进一步认识的动力。

不过还有更重要的第三个特征。一个好的统一理论应该提出前所未有的预言。它甚至可能提出只有在新理论的观点下才有意义的新型实验。最重要的还是，预言必须经过实验证明。

如果要判断眼下的统一研究的前景，我们就要寻求这样的三个准则——惊奇、新认识和实验证明的新预言。

物理学家深感统一的急迫，甚至有人说，向统一迈出的任何一步，都是向真理靠近的一步。但生活并不那么简单。任何时候都会有不止一条道路能统一我们所知道的事物——它们沿着不同的方向指引着我们的科学。16世纪，我们有两个截然不同的统一思路。一个是古老的亚里士多德和托勒密的理论，根据这个理论，行星与太阳和月亮统一为天球的不同部分。而另一个是哥白尼提出的新观点，将行星与地球统一到太阳的周围。两种思想都对科学产生了重大影响，但最多只能有一个是正确的。

从这儿我们可以看到选择错误统一的代价。假如地球在宇宙中心，那将极大影响我们对运动的认识。在天空中，行星改变方向是因为它们在确定的圆形轨道上不停地旋转。地球上的事物不会发生这样的事情：我们推动或扔出的东西，最终都会静止下来。地球上的物体不在天体的圆周轨道上运行，当然处于那样的状态。因此，在托勒密和亚里士多德的宇宙中，静止与运动有着分明的界线。

在他们的世界里，天与地也有着极大的分别——地球上的事物与天空的事物遵从不同的法则。托勒密提出天空的特定天体，如太阳、月亮和五个已知的行星，在圆形轨道上运动，而那些轨道本身也沿着圆周运动。这些所谓的本轮能预言日食、月食和行星的运动——预言的精度达到千分之一，证明了太阳、月亮和行星的统一是多么富有成效。亚里士多德为地球在宇宙中心提出了一个自然的解释：它由地球的材料组成，而那些材料的自然状态不是圆周运动，而是寻找中心。

在这种观点的教育下成长起来的人，熟悉它对我们周围的事物有多么强大的解释能力。当他们看到哥白尼将行星当做地球而不是太阳一样的物体，一定会深感不安。如果说地球是颗行星，那么它和它上面的所有东西都处于不停的运动。那怎么可能呢？这违背了亚里士多德的定律：任何不在天体圆形轨道的事物都必然趋于静止。它也违背了我们的经验：假如地球在运动，我们怎么没有感觉呢？

这个困惑的答案在于科学中的一个最伟大统一：运动与静止的统一。这是伽利略提出的，后来成为牛顿的第一运动定律，也叫惯性原理：处于静止或匀速运动的物体，在不受外力干扰时，将保持静止或匀速运动的状态。

牛顿所谓匀速运动的意思是，在某个方向以一定速度的运动。静止不过是匀速运动的一个特例——它恰好以零速度运动。

运动与静止怎么可能没有区别呢？关键的一点是认识到，物体是否运动，没有绝对的意义。只有相对于某个观察者才能确定运动，而观察者可以动，也可以不动。如果你以不变的速度从我面前经过，那么，在我看来静止在桌上的一杯咖啡，对你来说就是运动的。

可是，难道观察者也不能说明他是否在运动吗？对亚里士多德来说，答案显然是肯定的。伽利略和牛顿只能回答不。如果说地球在运动而我们没有感觉，那一定是因为以不变速度运动的观察者不会察觉其运动的任何效应。因此，我们不能说自己是否在运动，运动必须定义为一个纯粹相对的物理量。

需要说明的是，我们在谈匀速运动——在直线上的运动。（虽然地球当然不是沿直线运动，但偏离很小，不能直接觉察。）当我们改变速度的大小或方向，我们就有感觉。这样的改变就是我们所谓的加速度，加速度可以有绝对的意义。

伽利略和牛顿赢得了一场漂亮而微妙的理性的胜利。对其他人来说，运动与静止显然是截然不同的两个现象，很容易区分开来。但惯性原理将两者统一起来了。为了解释它们为什么看起来那么不同，伽利略提出了相对性原理。这个原理告诉我们，运动与静止的区别只有相对于观察者才有意义。因为不同观察者以不同方式运动，他们对物体运动或静止的判断也不同。于是，不同观察者的区别依然存在，这是当然的。因此，物体是否运动不再是一个需要解释的现象。在亚里士多德看来，任何物体如果在运动，一定有力作用在它上面。在牛顿看来，如果运动是匀速的，则它将持续下去，而不需要力的作用。

这是后来理论反复引用的有力论证。为了统一看似不同的事物，一种办法就是证明那种表面的差别源于观察者的不同观点。以前认为绝对的差别现在就成为相对的。这种统一方式很少见，代表着最高形式的科学创造力。它的实现将极大改变我们的世界观。

说两个看似截然不同的事物是相同的，通常需要大量的解释。只有某些情形，我们才能侥幸地将表面的差异解释为不同观点的结果。更多的情形，我们想统一的两件事本来就是不同的。这时候，为了解释看似不同的事物在某种意义上是相同的，理论家们可能陷入很大的困境。

我们来看看认为恒星就是太阳的布鲁诺的结果。恒星看起来比太阳暗淡得多，如果它们和太阳一样，就一定距离我们很远。布鲁诺需要的距离远远超过了那时人们相信的宇宙的大小，因此他的思想一开始就显得荒谬。

当然，这是提出新预言的机会：假如你能测量到恒星的距离，你就能发现它们比行星远得多。如果布鲁诺时代能做这样的测量，他大概能避免火刑的厄运。但要几百年后，人们才能测量恒星的距离。从实践的角度说，布鲁诺所做的，在当时技术条件下，是无法验证的断言。布鲁诺随意地将恒星推到那么远的距离，当然没人能检验他的思

想。

因此，为了解释如何统一不同的事物，有时我们会被迫面临新的完全不可能检验的假设。正如我们看到的，这并不意味着我们错了，但它也真实地说明了新统一的创立者们很容易陷入危险的境地。

事情也可能更糟。这些假设常常令它们自己更加混乱。实际上，哥白尼需要假定恒星很远。如果恒星像亚里士多德想的那么近，你可能会否定地球的运动——因为地球运动时，恒星的相对视位置会发生改变。为了解释为什么没有看到这样的效应，哥白尼和他的追随者们只能相信恒星非常遥远。（当然，我们现在知道恒星也在运动，不过因为距离太远，它们在天空的位置只有极其缓慢的变化。）

但假如恒星真有那么远，我们怎么能看到它们呢？它们一定很亮，也许像太阳那么亮。于是，布鲁诺提出的充满着无限多恒星的宇宙自然满足哥白尼的地球和行星一样运动的思想。

我们从这儿看到不同的统一思想常常会殊途同归。恒星与太阳统一，行星与地球统一，是和谐相容的两个思想，它们都要求运动与静止是统一的。

这些在16世纪的新奇思想与众多的其他思想相矛盾。托勒密关于行星与太阳和月亮一样都在本轮上运动的思想，迎合了亚里士多德的运动理论，统一了地球上的所有已知现象。

于是我们最终得到两组思想，每一组都由几个统一纲领构成。因此，危险的是整套的思想，其中不同的事物统一在不同的层次。在争论未决之前，相信任何一边都是合理的，都可能得到观测的支持。有时，甚至同一个实验可以解释为几个相互竞争的统一理论的证据。

怎么会这样呢？我们考虑一只从塔顶落下的球。结果呢？球落向地面，砸在塔基，而没有向西方偏离。好了，你可以说，哥白尼和他

的追随者们显然错了，因为这证明地球没有绕着自己的轴旋转。如果地球在转动，球应该落在远离塔基的地方。

但伽利略和牛顿统一可以宣称下落的球证明了他们的理论。惯性原理告诉我们，如果球在放下时随地球向东运动，那么它在下落过程中将继续那样运动。但球与塔是以相同速度向东运动的，所以它落在塔基。同样一个证据，亚里士多德哲学家可以用它来证明伽利略错了，而伽利略却拿来证明自己的理论是正确的。

那么，我们究竟如何确定哪个统一正确，哪个统一错误呢？有时候，某一点的证据会占尽优势，一个假设会远比其他假设更有成效，因而凡是理性的人都别无选择，只能赞同那个假设被证明了。就牛顿革命而言，最终出现了真正的地球运动的证据，不过在那证据之前，牛顿定律已经在诸多方面得到了证明，因而不可能逆转。

然而，在科学革命的进程中，相互竞争的假设经常都能找到合理的证据。我们现在就处于这样的时期，在以下的章节里我们就会考察那些对立的统一思想。我将尽量解释支持不同方面的证据，同时也要说明为什么科学家还没有达成共识。

当然，我们要多加小心。并非所有支持某个观点的证据都有坚实的基础。有时人们说支持某个困境中的理论，只不过是为自己找理由。最近我从伦敦到多伦多的飞机上碰到一伙人站在走廊里。他们向我打招呼，问我从哪儿来。我告诉他们我刚参加了一个宇宙学会议，他们马上问我如何看进化论。“噢，不，”我想了想，接着告诉他们自然选择已经证明是正确的，毫无疑问。他们自我介绍说是圣经学会的会员，刚从非洲回来。看来，他们去非洲的目的之一是为了检验特创论的某些原理。他们想拉我一起讨论，我警告说他们会输的，因为我有很好的证据。“不会的，”他们坚持说，“你并不了解事实的全部。”于是我们展开了论战。我说，“但你们当然接受这样一个事实：我们有很多已经灭绝了的生物的化石。”他们回答，“不！”

“你们说不，是什么意思？那恐龙呢？”

“恐龙还活着，在地球上游荡呢！”

“真荒唐！在哪儿呢？”

“非洲。”

“非洲？非洲到处是人。恐龙那么大，怎么没人看见过呢？”

“它们生活在密密的丛林里。”

“那一定有人见过它的。你是不是要说你认识某个见过它的人？”

“俾格米人（pygmy）告诉我们，他们偶尔看见过恐龙。我们找了，没找到，但我们在树干上18到20英尺的高度看到了它们留下的痕迹。”

“所以你们认为那是大动物留下的。可化石证据表明它们是成群生活在一起的。为什么只有那些俾格米人见过呢？”

“原因很简单。它们大多数时间都在洞穴里冬眠。”

“在丛林里？丛林里有洞穴吗？”

“有啊，当然了，为什么没有呢？”

“能让大恐龙进入的洞穴？如果洞穴那么大，应该很容易找到的，那你们就可以往里面看，看见它们在睡觉。”

“为了保护自己，那些恐龙在冬眠时会用泥土把洞口都封起来，这样就没人知道它们在那儿了。”

“它们怎么能把洞穴封得那么严实而不让人看见？它们是用爪子还是鼻子来运泥土呢？”

这时，特创论者们承认他们不知道，但他们告诉我，他们学派的“圣经生物学家”正在丛林里找恐龙呢。

“如果他们找到了活恐龙，一定要告诉我。”我说，然后走回自己的座位。

我没有虚构故事，讲这个故事也不是为了让大家好玩儿。它说明理性并不总是简单的练习。有的理论预言了我们永远没见过的东西，不相信那样的理论，通常是很有道理的。但有时看不见的东西也有很好的理由。毕竟，如果真有恐龙，那它们一定藏在某个地方。为什么不能在非洲丛林的洞穴里呢？

这也许显得愚昧，但粒子物理学家们不止一次地感到，为了让某个理论或数学的结果有意义，他们必须构造看不见的粒子，例如中微子。为了解释为什么中微子难以探测，他们只好让中微子的作用很微弱。在这个例子中，这种策略是正确的，因为多年以后，人们设计了寻找中微子的实验。它们的作用的确很微弱。

所以，有时候，一个好理论即使预言了没见过的东西，也有理由把它留下来。有时候，我们被迫做出的假设后来证明是正确的。提出这样特殊的假设，不仅使思想合理，有时也能预言新的现象。但有时候，我们也可能犯轻信的毛病。从这点说，穴居恐龙也许有道理。当我们面临一个过去的好思想可能会变得毫无价值时，首要的问题是判断。在训练有素的聪明人之间，肯定有众说纷纭的情形。但最终会有证据说话的时候，那时任何有理性的公正的人都不会再认为那个思想有什么道理。

要判断我们是否到了那样的转折点，方法之一是考虑其唯一性。

在科学革命中，任何时候都有几个不同的统一思想，可能把科学引上矛盾的方向。这是正常的，而且在革命进程中也没有什么合理的根据来选择具体的哪一个。在这种时候，即使有很聪明的人匆匆做了选择，常常也可能是错的。

不过最终会有某个思想比其他思想能解释更多的东西，而它通常是最简单的。在这种情形，一个思想在产生新思想、满足实验、解释能力和简单性等方面远远超越了其他思想，它就具备了唯一性。我们就说它有真理的特征。

怎么会这样呢？我们考虑德国天文学家开普勒（Johannes Kepler，1571～1630）一个人提出的三种统一思想。开普勒一生对行星着魔，因为他相信地球是行星，他知道从水星到土星的六颗行星。它们的运动已经观测了几千年，有大量数据。最精确的数据来自丹麦天文学家第谷（Tycho　Brahe）。开普勒为了得到那些数据，最终来到第谷手下工作（第谷死后，他把数据偷走了，不过那是另一个故事了）。

每个行星轨道有一个半径，每颗行星有一个轨道速度。另外，行星速度不是均匀的，在环绕太阳的过程中时快时慢。这些数据看起来杂乱无章。开普勒花了一生的精力来寻求一个能统一行星运动的原理，从而用那个原理来解释行星轨道的数据。

起初，开普勒根据古老的传统来统一行星，认为宇宙理论只能用最简单的几何图形。古希腊人之所以相信圆周在圆周上运动的图景，是因为圆是最简单的闭合曲线，因而在他们看来也是最美的。开普勒想寻求同样美妙的几何图形来解释行星轨道的大小。他发现了一个非常精美的思想，如图2-1。

假定我们先知道了地球轨道，那么接下来需要解释五个数，即其他5颗行星的半径与地球半径之比。如果能找到一个解释，必然存在

某个美妙的几何结构正好能给出那五个数，不多也不少。那么，是否有那么一个几何问题刚好有5个答案呢？

是的。立方体是一种完美固体，它的每个面都一样，每个边也一样长。这样的固体叫柏拉图固体。有多少种呢？正好5种：除了立方体，还有正四面体、正八面体、正十二面体和正二十面体。

开普勒没费多少时间就做出了一个有趣的发现。将地球轨道嵌入球内，球外接一个正十二面体。在它们外面嵌一个球，火星轨道就在那个球面上。外接一个正四面体，四面体外接另一个球，则木星轨道就在那个球面上。木星轨道外接立方体，土星就在它外面。在地球轨道内，开普勒内接正二十面体，金星就环绕着它，金星轨道内接正八面体，就是水星。

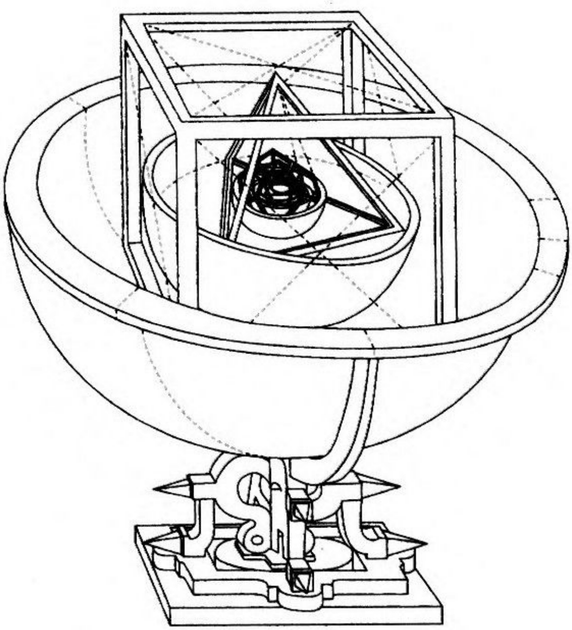


图2-1　以柏拉图固体为基础的开普勒的第一个太阳系理论

这个统一理论解释了行星的轨道半径，以前还没有理论这样做过。这是一个洋溢着数学美的理论。可为什么没人相信呢？虽然理论很动人，却没有引出什么东西。以它为基础没有预言任何新的现象，甚至它也不能使人们认识行星的轨道速度。这个思想太静态了，尽管统一，却没有将科学引向任何有趣的地方。

开普勒为此思考了很长时间。轨道直径解释了，他只需要解释不同行星的速度。最后他提出，行星在运行中“歌唱”，音调的频率正比于速度。不同行星在轨道运行，以六种声音唱出一曲和谐的歌，开普勒称那是天球的和谐。

这个思想也有古老的根源，令人想起毕达哥拉斯的发现：音乐的和谐源于简单的数字比例。不过它的问题也很明显。它不是唯一的：六种声音可以有多种和谐的方式。更严峻的是，后来发现行星不止6颗。而且，与开普勒同时的伽利略发现了木星的4颗卫星。所以，天上还有另一个轨道系统。如果开普勒的理论是正确的，它们也该适用于新发现的系统。可是它们不能。

除了这两个宇宙的数学结构，开普勒还做出了为科学带来实在进步的三个发现，那就是他多年分析偷来的第谷数据之后提出的著名的三大定律。这些发现一点儿也不如前两个思想那么优美，但它们很成功。而且，其中之一没有别的办法可以实现，那就是速度与轨道直径的关系。开普勒的三个定律不但满足所有六颗行星，也满足木星的卫星。

开普勒发现那些定律，是因为他将哥白尼的统一引向了逻辑的结果。哥白尼说过太阳处于（或邻近）宇宙的中心，但在他的理论中，行星的运动与是否有太阳无关。太阳的唯一作用就是照亮天空。哥白尼理论的成功启发开普勒提出这样的问题：太阳在行星轨道的中心附近，是否真的只是偶然？他想知道太阳是否可能在驱动行星轨道中起着某种作用。太阳会不会以某种方式将力作用在行星上，而那种力正

是行星运动的原因？

为了回答这些问题，开普勒必须为太阳在每个行星轨道的精确位置寻求一种作用。他的第一个突破是发现了轨道不是圆，而是椭圆。太阳也有了准确的位置：它恰好处于每个轨道椭圆的焦点。这是他的第一定律。不久之后，他发现了第二定律，即行星在轨道的速度随着接近或远离太阳而增大或减小。后来他又发现了第三定律，决定了行星的速度之间有什么关系。

这些定律指向太阳系背后的某个深层的事实，因为它们适用于所有行星。结果是我们第一次有了一个能做出预言的理论。假定发现了一颗新行星，我们能预言它的轨道吗？在开普勒之前，没人能做到。但有了开普勒定律，我们只需要观测它的位置的两点就能预言它的轨道。

这些发现为牛顿铺平了道路。正是牛顿的洞察力发现了太阳作用于行星的力与地球作用于我们的引力是同一种力，从而统一了天上与地上的物理学。

当然，对当时的大多数科学家来说，太阳向行星发出力的思想是很荒唐的。他们相信空间是虚空的，没有能传递那种力的介质。而且，它也没有可见的表现——没有从太阳伸向行星的臂膀——而看不见的东西不可能是真的。

想做统一的人可以从这儿得到几点很好的教训。一是数学美可能误导。数据的简单观察通常更为重要。另一点教训是，正确的统一理论能对当时确凿无疑的现象发生作用，例如开普勒定律对卫星的应用。正确的统一也可能引发当时看来荒唐的问题，但能导致进一步的统一，就像开普勒假定从太阳作用于行星的力。

最重要的是，我们看到真正的革命通常需要几个不同的统一思想

走到一起来相互支持。在牛顿革命中，有几个成功的统一：地球与行星的统一，太阳与恒星的统一，静止与运动的统一，地球引力与太阳对行星的作用的统一。单个地看，这些思想没有一个能流传下来；结合在一起，它们就所向无敌了。那结果就是一场彻底转变我们自然知识的大革命。

在物理学史上，有一个统一出类拔萃，成为物理学家在过去30年里追求的典范。那就是麦克斯韦在1860年代实现的电与磁的统一。麦克斯韦运用了强有力的场的概念——英国物理学家法拉第在19世纪40年代为解释力在虚空传播而提出的。概念的关键在于，场是一个物理量，像数一样，存在于空间的每一点。在空间运动时，场的数值连续改变。场在一点的值也随时间而变化。这个理论确立了场在时空中变化的规律。这些定律告诉我们，场在某点的值受它在附近其他点的值的影响。因此，场可以将力从一个物体带到另一个物体。我们再不需要相信什么可怕的超距作用了。

法拉第研究的一种场就是电场。它不是单纯的数，而是一个矢量，可以想象为一个箭头，它能改变方向，也能改变大小。想象空间每一点的这种箭头，相邻点的箭头的尖由橡皮筋连接。如果我们拉动一个箭头，它就会拉动相邻的箭头。箭头也受电荷影响。箭头在电荷影响下自我调整，由正电荷指向负电荷（图2-2）。

法拉第也研究过磁。他发明了另一种场（即另一种箭头的集合），他称之为磁场。这些箭头总是指向磁体的两极。

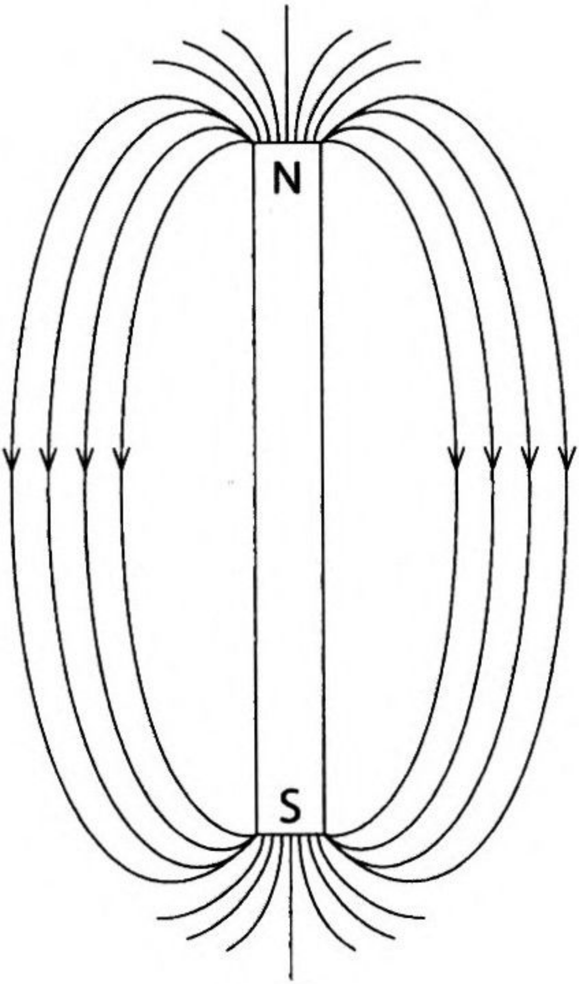



图2-2　代表磁铁棒磁场方向的磁力线

法拉第写出了几个简单法则，描述了电磁场如何受邻近电荷和磁极以及邻近场的矢量的影响。他和别人检验了这些法则，发现它们的预言和实验是一致的。

那时还发现了电磁混合的现象。例如，在圆周运动的电荷会产生磁场。麦克斯韦意识到这些发现意味着电与磁的统一。为了实现统一，他需要改变方程。为了改变方程，他只需要增加一项，于是他的

统一就成为有实际作用的统一。

新方程允许电磁场相互转化。在这种转化中，电场与磁场交互产生，从而形成一种在空间移动的波。前后舞动一个电荷，也能产生那样的波动模式。生成的波可以将能量从一个地方带到另一个地方。

最令人惊奇的是，麦克斯韦还能根据他的理论计算波的速度。他发现那速度和光速是一样的。接着他一定大吃一惊：在空间穿行的电磁波就是光。麦克斯韦没打算建立一个光的理论，而是想统一电与磁。但在统一的过程中，他实现了更重要的事情。这个例子说明，一个好的统一对理论和实验都会产生意外的结果。

新预言接踵而来。麦克斯韦意识到，应存在所有频率的电磁波，而不仅限于可见光，于是就发现了无线电波、紫外线、红外线等等。这儿还留下另一个历史教训：当新的正确的统一出现时，其意义会很快显露出来。许多这样的现象，都是在麦克斯韦理论发表后的几年内发现的。

这就提出一个问题，在我们讨论其他统一时可能变得很重要。所有统一都会有结果，因为统一的事物能相互转化，能导致一系列新现象的出现。如果运气好，很快就能观察到那些新现象——统一理论的创立者们当然有权利为自己欢呼。但我们将看到，在其他情形，预言的现象已经与观测结果有了矛盾。在这种不幸的情况下，理论家要么被迫放弃理论，要么人为地将它限制起来，隐藏那些统一的结果。

可是，麦克斯韦的电磁统一理论虽然成功了，却面临着一个难以逾越的障碍。19世纪中叶，多数物理学家相信物理学已经统一了，因为所有事物都是物质组成的（为了满足牛顿定律，它们也必须是物质组成的）。对这些“机械论者”来说，在空间波动的场的概念很难理解。在麦克斯韦理论中，电磁场外没有能让他们感觉实在的弯曲和伸展的东西，因而是没有意义的。当我们通过光看见花朵时，一定应该

有什么物质的东西在颤动。

法拉第和麦克斯韦都是机械论者，他们也费了很多时间和精力来回答这个问题。除了他们，许多著名研究机构的年轻人为了美好前程，也在为电磁场精心构造他们作为麦克斯韦方程基础的微观齿轮、滑轮和传送带。结果产生一些错综复杂的方程，谁能解那些方程，谁就能获奖。

这个问题有一点显著的表现，那就是，光从太阳和恒星来到我们，而外太空是空无一物的。假如空间有任何物质，它将阻碍行星的运动，那么行星早就落进太阳里了。可是，电场和磁场又怎么能在虚空里呢？

于是，机械论者发明了一种新的物质形式——以太，并将它充满空间。以太有着矛盾的性质：它必须极端致密而坚硬，因为光要像声波一样通过它；而光速与声速的巨大差别就是以太的超大密度的结果。同时，以太对穿过它的普通物质没有任何阻碍作用。这一点比看起来更难满足。我们只能说，以太与普通物质不发生相互作用——就是说，它们彼此没有力的作用。可是，如果光（或电磁场）只不过是以太的应力，普通物质为什么能探测到它们呢？难怪，谁能明白这些问题，就能做教授。

还会有比以太更优美的统一吗？它不仅统一了光、电和磁，还统一了物质。

然而，正当以太理论发展时，物理学家的物质概念也在发生改变。19世纪初，多数物理学家都认为物质是连续的，可是世纪之末，人们发现了电子，至少部分物理学家开始重视物质由电子组成的观点。但那就引出另一个问题：在以太的世界里，原子和电子是什么呢？

画出场的力线，像磁场的力线那样，从磁北极指向磁南极。场线的终点在磁极，其他地方不会中断。这是麦克斯韦的一个定律。但场线可以形成闭合线圈，线圈可以自己形成结。所以，原子也许是磁力线的结。

可是，正如水手都知道的，打结有不同的方式。那样也许正好，因为有不同的原子。于是，剑桥的一个著名教授提出，不同原子对应于不同的结。

这看起来很荒唐，但想想那个年代，19世纪90年代和20世纪初，我们对原子懂得很少。我们那时还不知道原子核，没听说过质子和中子。所以，这样的思想算不得疯狂。

那时候，我们对线圈的结也知道得很少。没人知道打一个结有多少方式，又如何区分它们。于是，在这种思想启发下，数学家开始研究如何区分各种可能的结。这慢慢演进为一个叫纽结理论的数学领域。很快证明，打一个结有无穷多种不同的方式，但过了很久人们才发现如何区分它们。20世纪80年代有了一些进展，但仍然不知道以什么过程来判断两个复杂的结是相同还是不同。

我们看到，一个好的统一思想，即使证明是错的，也能激发新的追寻的路线。然而，我们应该记住，仅仅因为统一理论结出了数学成果，并不能说明那个物理理论是正确的。相反，结理论的成功仍然要求我们相信原子是磁场里的结。（不过，正如我们将在第15章看到的，也许这并不完全是错的。）

还有一个问题：麦克斯韦理论似乎与牛顿物理学的相对性原理相矛盾。结果证明，研究电磁场的观察者可以通过各种实验（包括测量光速）来判断他是否在运动。

还有一个矛盾存在于两个统一之间，而那两个统一都是牛顿物理

学的核心：服从牛顿定律的物质的统一与运动和静止的统一。对大多数物理学家来说，答案是显然的：物质宇宙的观念当然更重要，而运动难以确定，也许只是微不足道的事情。但也有少数人认为相对性原理才是更重要的问题。那些人中间，有个年轻的学生，在苏黎世读书，他的名字叫爱因斯坦。他为这个问题沉思了10年，从16岁开始，最后在1905年他意识到问题的解决需要彻底改变我们对空间和时间的认识。

爱因斯坦解决问题的方法，就是牛顿和伽利略在建立运动的相对性原理时用过的技巧。他认识到电效应与磁效应的区别依赖于观察者的运动。所以，麦克斯韦的统一比他本人原来想象的更加深刻。电场与磁场不仅是同一个现象的不同方面，不同的观察者也能做出不同的判断。就是说，一个观察者可以用电来解释某个特殊现象，而另一个相对于他运动的观察者可以用磁来解释那个现象。但两个观察者对发生的事件有一致的看法。就这样，爱因斯坦的狭义相对论诞生了，它结合了伽利略的静止和运动的统一与麦克斯韦的电和磁的统一。

理论产生了很多结果。一个结果是光必然具有普适的速度，与观察者的运动无关。另一个结果是，空间与时间必然是统一的。从前，二者截然不同：时间是普适的，对两个同时发生的事件，每个人都会做出相同的判断。爱因斯坦证明，两个事件是否同时发生，相对运动的两个观察者有着不同的看法。时空的统一隐含在他1905年的题为“论运动物体的电动力学”的论文里，他的老师闵可夫斯基（Hermann Minkowski）在1907年将它明确表达了出来。

于是，我们又看到了两个在竞争的统一。机械论者有一个优美的统一物理学的思想：万物都是物质的。爱因斯坦则相信另一种统一，即运动与静止的统一。为了支持那个观点，他不得不发展一个更深层的统一——空间与时间的统一。不论哪种情形，过去认为绝对不同的东西只有相对于观察者的运动才会不同。

最后，两个统一的矛盾由实验解决了。如果你相信机械论者，你会相信观察者能够测量他穿过以太的速度。如果你相信爱因斯坦，你就知道他做不到，因为所有观察者都是平等的。

在爱因斯坦提出狭义相对论的1905年之前，人们几度尝试探测地球在以太中的运动，都失败了。[10]以太理论的支持者们调整了他们的预言，结果只是使探测地球的运动变得越来越艰难。这是很容易发生的，因为他们计算时用的是麦克斯韦理论，而正确解读那些方程，是跟爱因斯坦的预言一致的，即不可能探测出那样的运动。就是说，机械论者有了正确的方程，但是解读错了。

至于爱因斯坦本人，我们还不清楚他对先前的实验了解多少，但它们不会有多大作用，因为他已经相信地球运动是不可能探测的。爱因斯坦还只是刚开了头。正如我们将在下一章看到的，空间和时间的统一将走得更远、更深。当多数物理学家都跟上来接受狭义相对论时，爱因斯坦已经远远走到前头了。

## 第三章　几何世界

20世纪初的几十年里，人们进行过几次统一的努力。有几个成功了，其他的都失败了。简单回顾那些故事，从中汲取教训，能帮助我们认识当前的统一所面临的危机。

从牛顿到爱因斯坦，有一个思想一直占着统治地位：世界是物质组成的，而没有别的东西。连电与磁也是物质的不同方面——它们不过是以太的应力而已。但当狭义相对论胜利时，这幅美妙的图像就破灭了。因为，假如静止与运动的整个概念都失去了意义，以太必然就是虚幻的东西。

统一要在其他地方去寻找，而确实有那么一个地方。那就是颠倒以太理论：如果场不是物质组成的，那么场也许就是那基本的东西，

因而物质必然是场构成的。那时已经有了电子和原子作为场的应力的模型，所以迈出这一步并不困难。

但是，虽然这个思想赢得了信徒，疑惑仍然存在。例如，有两个不同的场，引力场和电磁场。为什么是两个而不是一个呢？这就到头了吗？对统一的渴望激励着物理学家们问：引力场与电磁场是不是同一现象的两个方面？于是出现了我们今天说的统一场论。

因为爱因斯坦只是把电磁场纳入了他的狭义相对论，所以，最合乎逻辑的路线是修正牛顿的引力理论，使它与狭义相对论协调。这是容易实现的。不仅如此，牛顿理论的修正还引出一个精彩的发现，成为今日统一理论的核心。1914年，一个叫诺德斯特罗姆（Gunnar Nordstrom）的芬兰物理学家发现，为了统一引力场与电磁场，我们只需要增加一个空间维度就行了。他写出描述有4个空间维（加1个时间维）的世界的电磁场的方程，引力也就跳出来了。正是靠这额外的1个空间维，我们就得到与爱因斯坦狭义相对论相容的引力与电磁力的统一。

但假如真是那样，那我们在遥望三维空间时，岂不应该看到那个新的维度吗？假如不是那样，这个理论岂非明显错了？为了避免这个麻烦的问题，我们可以让那个新维度是一个圆圈，这样，当我们看到它时，其实是绕了一圈又回到同一个地方。[11]于是我们可以令圆半径很小，因而很难发现还有那么一个多余的维。为什么收缩一个东西就能使它变得看不见呢？回想一下，光是波组成的，每个光波有一定波长（即两个相邻波峰间的距离）。光波的波长决定了我们所能看见的最小尺度，因为我们不能确定比我们所用的波长更小的物体。维度缩小的唯一效应就是一种具有引力的一切性质的新力。

大家也许认为爱因斯坦会张开怀抱接受这个新理论。但那时（1914年）他已经走上了一条不同的道路。他选择了一条与众不同的统一引力与相对论的路线，将他带回到相对性原理的基础：伽利略在

几个世纪前发现的运动与静止的统一。那个统一只涉及了匀速运动——即以不变速度在直线上的运动。1907年初，爱因斯坦开始为自己提出其他运动形式的问题，如加速运动，即大小或方向变化的运动。加速运动与非加速运动的区别是不是也应该以某种方式消除呢？

乍看起来，这一步似乎迈错了。因为，虽然我们不能感觉匀速运动的效应，但我们实在体会了加速度的效应。当飞机起飞时，我们感觉有什么在把我们推回座位。当电梯开始上升时，我们会感觉加速度以额外压力的形式将我们推向地板。

就在这一点上，爱因斯坦发挥了他非凡的洞察力。他意识到加速度效应与引力效应是无法分辨的。想象一个妇人站在电梯等着它启动。这时，她已经能感觉有一个力将她拉向地板。当电梯开始升起时，事情没有本质的不同，只是程度的差异：她感觉同一个力在增大。假定电梯处于静止，但引力的强度在瞬间增大，情况如何呢？爱因斯坦发现，她的感觉就像电梯突然加速上升一样。

我们还可以反过来看这一点。假定拉电梯的钢索被割断了，电梯和里面的乘客开始下落。在自由下落中，乘客会感觉失重，和宇航员在飞船轨道上的感觉一样。就是说，下落电梯的加速度完全抵消了引力的效应。

爱因斯坦后来回忆，他认识到从楼顶落下的人在下落过程中不会感觉到引力的作用。他说这是他“一生最幸运的思想”，并使其成为原理，即他所谓的等效原理。原理说，加速度效应不可能与引力效应区别。12

于是，爱因斯坦成功统一了所有类型的运动。匀速运动与静止没有区别，加速度与静止也没有区别，不过引力场出现了。

加速度与引力的统一有着重要的意义。即使在它的基本意义清楚

之前，也有巨大的实验意义。甚至只用高中代数就能根据它做出某些预言——例如，时钟在引力场中变慢，最终也的确证实了。另一个预言（最早是爱因斯坦1911年提出的）是，光线在经过引力场时会发生弯曲。

注意，和以前讨论过的成功统一理论一样，同时出现了更多的统一。因为两种运动都统一了，所以不再需要区分匀速运动与加速运动。加速运动的效应与引力统一了。

即使爱因斯坦可以根据等效原理导出几个预言，新原理还算不上一个完整的理论。建立一个完整的理论，是他一生最大的挑战，耗费了他近10年的时间。为什么呢？我们先来看看引力弯曲光是什么意思。在爱因斯坦的这个特别观点出现之前，世界上存在两类不同的东西：空间里的万物和空间本身。

我们不习惯把空间看作具有自身特点的实体，但它确实是那样的。空间有三维，也有一种特殊的我们在中学学过的几何，即所谓的欧几里得几何——2000多年前的欧几里得建立了它的公设和定理。欧几里得几何就研究空间自身性质，它的定理告诉我们空间的三角形、圆和直线是怎么样的。但它们只对具体的对象成立，不论物质的对象还是想象的对象。

麦克斯韦电磁理论的一个结果是，光沿直线运动。因此我们有理由用光线来探究空间的几何。但如果我们接受这个思想，立刻会看到爱因斯坦的理论有着重大意义。因为光线被引力场弯曲，而引力场源于物质的存在，于是我们能得到的唯一结论是，物质的存在影响空间的几何。

在欧几里得几何中，如果直线原先是平行的，则它们永远不会相交。但原先平行的两条光线却可能在真实世界里相交。因为，它们在经过一颗恒星的两边时，会向对方弯曲。所以，爱因斯坦几何不是现

实世界的几何。而且，那个几何在不停地变化，因为物质在不停地运动。空间几何不是平直的、无限的平面，而是像海洋的表面那样——不息地动荡，起伏着大大小小的波涛。

于是，空间几何原不过是另一个场。其实，空间几何与引力场几乎是一样的。为了解释这一点，我们需要回想爱因斯坦在狭义相对论里实现的空间与时间的部分统一。在那个统一里，空间和时间一起形成一个四维的叫时空的实体。从以下的角度看，它具有类似于欧几里得几何的几何。

考虑空间的一条直线，两个粒子沿着它运动，但一个匀速，另一个不停地加速。就空间来说，两个粒子在同一条路线上运动。但它们在时空中走着不同的路线。速度不变的粒子，不仅在空间而且在时空里都沿直线运动，加速粒子则在时空中沿曲线运动（图3-1）。

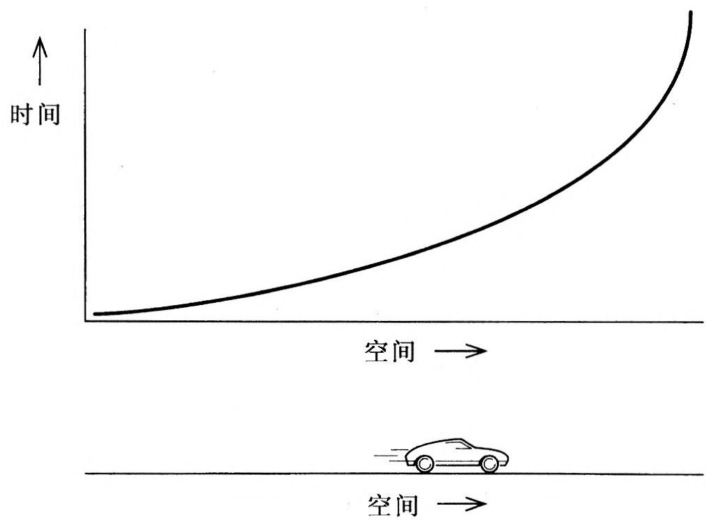



图3-1　在空间直线上减速运动的小汽车，在时空中沿曲线运动

于是，正如空间几何能区分直线和曲线，时空几何能区分匀速运动与加速运动。

可是，爱因斯坦的等效原理告诉我们，在小距离范围内，引力效应不可能与加速效应区分。[13]而时空几何能告诉我们哪些轨道是加速的，哪些是没有加速的，因而它描述了引力效应。于是，时空几何是引力场。

这样，等效原理的两重统一成了三重统一：考虑引力效应后，所有运动都是等价的；引力与加速度不可区分；引力场与时空几何统一。当其中的细节清楚之后，它们就成为爱因斯坦在1915年最终形成的广义相对论。

对开始没找到研究工作的年轻人来说，这是很幸运的。

于是，物理学在1916年面临着两个截然不同的未来，其基础都是统一引力与其他物理学。一个是牛顿的引力与电磁力的优美统一，它只是简单地添加了一个隐藏的空间维；另一个就是爱因斯坦的广义相对论。两个似乎都是和谐的，而且都有意外的成果。

它们不可能都是正确的，我们必须进行选择。幸运的是，两个理论都有可以实验的预言。爱因斯坦的广义相对论预言引力必然使光线弯曲——而且精确预言了弯曲多少。在牛顿理论中则没有这种效应：光总是周期性地走直线。

1919年，英国大天体物理学家爱丁顿（Arthur Eddington）带领一个探险队远征非洲西海岸，观测当年的日全食，证明了太阳的引力场确实弯曲了光线。因为日食发生的时候，可以看到太阳的边缘附近的星光，而那些恒星本来是在太阳的背后。假如太阳的引力场没有弯曲星光，那些恒星是看不见的。但我们看见了。这样，我们通过唯一可能的方式，即通过实验，对两个截然不同的统一方向做出了选择。

这是一个重要事例，因为它说明了我们单靠思想能走多远。有些物理学家认为，广义相对论的例子证明了纯思想足以指引前进的方

向。但事实恰恰相反。假如没有实验，多数物理学家可能会选择诺德斯特罗姆的统一，因为它很简单，而且带来了强有力的新思想，通过额外空间维实现统一。

爱因斯坦对引力场与时空几何的统一，标志着我们对自然的思维方式的深刻转变。在爱因斯坦之前，人们认为空间和时间具有一些固定在所有实体的性质：不论过去、现在还是将来，空间几何都是欧几里得几何。事物在空间里运动，在时间中演化，但空间和时间本身永远不变。

对牛顿来说，空间和时间构成绝对的背景，它们为自然的大戏提供固定的舞台。空间和时间的几何只是用来描述变化的事物（如粒子的位置和运动），但它们本身从不改变。我们为依赖于这种绝对固定框架的物理学理论起了一个名字，称它们是背景相关理论。

爱因斯坦的广义相对论完全不同。它没有固定的背景框架。空间和时间的几何与自然的一切事物一样，也变化和演进。不同的时空几何描述不同宇宙的历史。我们不再有在固定背景几何下运动的场。我们有很多相互作用的场，都是动态的，都相互影响，其中之一就是时空的几何。我们称这样的理论为背景独立理论。

记住背景相关和背景独立理论的特点。随着本书的展开，我们会看清它们之间的区别。

爱因斯坦的广义相对论经受了我们在上一章讲的那些检验统一理论的所有考验。它的统一隐含着一些概念的变革，很快导致了新的预言，如膨胀宇宙、大爆炸、引力波和黑洞，所有这些预言都有很好的证据。我们的整个宇宙学观念都以它为基础。诸如光被物质所弯曲等等看似极端的观念，现在已经成为我们探究宇宙物质分布的工具。理论的预言在每一次接受细节的检验时，都漂亮地通过了。[14]

但广义相对论只是一个开始。甚至在爱因斯坦发表最后形式的理论之前，他和别人就已经在建立新的统一理论了。他们有一个共同的简单想法：如果引力可以理解为空间几何的外在表现，那电磁力为什么不能也如此呢？1915年，爱因斯坦写信给希尔伯特（David Hilbert，也许是当时健在的最伟大的数学家），“我常常苦思冥想，在引力与电磁力之间搭建一座桥梁。”[15]

但直到1918年才真正出现这种特殊统一的一个好想法。数学家外尔（Herman Weyl）创立的那个理论包含着美妙的数学思想，后来成为粒子物理学的标准模型的核心。可是它失败了，因为在外尔原来的理论中，有一些重要结论与实验相冲突。一个是物体的长度依赖于它的运动路径。假如你拿两根米尺，将其分开，然后再放到一起比较，一般会发现它们有不同的长度。这是比相对论还令人惊骇的结果。在相对论中，米尺确实也会显得不同，但那只发生在它们相对运动的时候，而不是在静止比较的时候。当然，它也和我们的自然经历相矛盾。

爱因斯坦不相信外尔的理论，但很欣赏它，写信给外尔说，“除了与相对论不一致，它无论如何算得一个美妙的智力演绎。”[16]外尔的回答说明了数学美的魔力：“您对我理论的拒绝令我感到很沉重……不过我本人的头脑仍然相信它。”[17]

不过，外尔第一次统一的尝试纵然是失败了，但他创立了统一的现代概念，最终导致了弦理论。他第一个（但肯定不是最后一个）宣称，“我想冒昧地说，我相信整个物理现象可以从一个普遍的具有最大数学简单性的自然法则推导出来。”[18]

外尔理论提出一年后，一个叫卡鲁扎（Theodor Kaluza）的德国物理学家发现，如果复活诺德斯特罗姆的隐藏维的思想，可以用不同的方法来统一引力与电磁力。可是他走了一条弯路。诺德斯特罗姆是通过将麦克斯韦的电磁理论用于五维世界（4个空间维和1个时间维）

而建立引力理论的。卡鲁扎的路线正好相反：他将爱因斯坦的广义相对论用于五维世界，然后建立电磁力。

我们可以在寻常三维空间的每一点放一个小圆圈（见图3-2），来感觉那种新空间的形式。新几何能以新的方式产生弯曲，因为不同点的小圆圈能以不同方式生成。于是，对原来三维空间每一点的度量来说，出现了新的东西。这个信息看起来恰好就像电场和磁场。

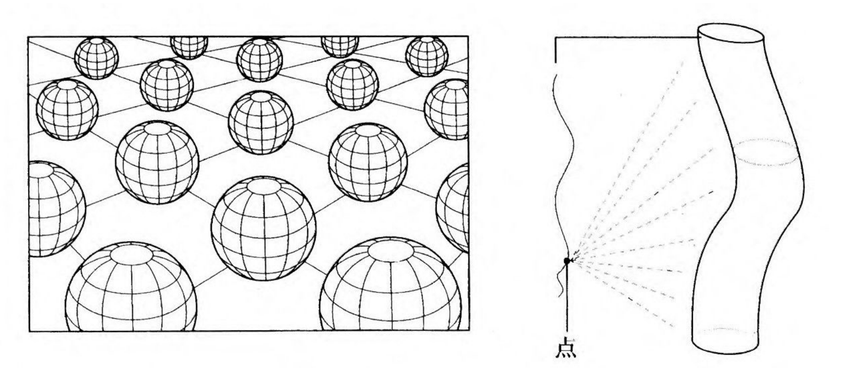



图3-2　卡鲁扎-克莱因理论用的卷曲的额外空间维。左：原来三维空间的每一点放一个球面，形成五维空间。右：一维空间上放一个小圆圈。从远处看，空间向一维的，但是在近处看，它是二维的

另一个精彩的结果是，电子的电荷关系着小圆圈的半径。这并不奇怪：如果说电场恰好是几何的表现，那么电荷也应该是。

还不仅如此。广义相对论以一组方程（叫爱因斯坦方程）描述了时空的几何。我不必把方程写出来，关键的一点在于，同样的方程也可以用于我们刚才讲的五维世界。只要我们加一个简单的条件，就可以看到它们能正确描述电磁场和引力，将它们统一起来。于是，如果这个理论是对的，那么电磁场不过就是五维几何的另一个名字。

瑞典物理学家克莱因（Oskar　Klein）在20世纪20年代重新发现并进一步发展了卡鲁扎的思想。他们的理论的确优美动人。只需要给

空间增加一个维，引力与电磁力就一下子统一了，而麦克斯韦方程便可解释为爱因斯坦方程的一个结果。

这一次，爱因斯坦着迷了。1919年4月，他写信给卡鲁扎说，“用五维圆柱来实现[统一]的思想，我从来没有想到过……乍看起来我非常欣赏你的想法。”[19]若干年后，在给荷兰物理学家洛伦兹（Hendrik Lorentz）的信中，他高兴地说，“看来，引力与麦克斯韦理论的统一已经通过五维理论而圆满地实现了。”[20]著名物理学家乌伦贝克（George Uhlenbeck）曾回忆，他在1926年第一次听到克莱因的思想时，“感觉入神了！现在总算有人理解我们的世界了。”[21]

遗憾的是，爱因斯坦和其他的热情者们都错了。和诺德斯特罗姆理论一样，通过加一个隐藏维来实现统一的思想失败了。我们有必要知道为什么。

我前面说过，一个成功的统一思想，必须用实验验证的新预言来确立其地位。成功的统一还会产生许多引出进一步发现的新思想。卡鲁扎一克莱因理论尽管迷倒了一些人，但它并不具备那些条件。前面说过，这个理论附加了一个额外的条件，要求多出的那个空间维卷曲成看不见的小圆圈。不但如此，为了从理论导出电磁力，圆圈的半径还必须是固定的，在空间和时间上都不发生改变。

这是整个理论的致命弱点，直接导致了它的失败。原因在于固定额外维的半径会破坏爱因斯坦广义相对论的实质，即时空几何是动力学的。如果我们另加一个广义相对论描述的空间维，那个维的几何也会成为动力学的。实际上，只要允许小圆圈的半径自由变化，情形就是那样。这样的话，卡鲁扎和克莱因的理论将有无穷多个解，其中的圆半径都能在空间和时间中变化。这蕴涵着奇妙的现象，因为它导致引力效应与电效应相互转化的过程。它还将导致电荷随时间而变化的过程。

但是，如果卡鲁扎-克莱因理论确实是统一的，第五维就不应该与其他维有什么不同：小圆圈必须是可以变化的。这样，它产生的过程就是统一电与引力的必然结果。如果看到了那些过程，那就直接证明了几何、引力、电和磁都是同一个现象的不同方面。遗憾的是，我们还没观察到那些效应。

有时候，理论家们可以很快欢呼统一的结果，但现在的情形不是这样的。相反，理论家们必须把统一藏起来，而只关注解的一个无穷小的部分——其中第五维的半径在空间和时间都是固定不变的。

情况比这更坏。原来，这样的解是不稳定的。如果稍微触动几何，小圆圈将立刻坍缩为标志时间终结的奇点。如果换一种方式触动几何，则小圆圈可能长大，额外的维会立刻变得可见，理论也就彻底破产了。于是，他们只好把理论的预言藏起来，掩盖它可能犯那么严重错误的事实。

在这一点上，爱因斯坦也失去热情了。他写信给朋友艾伦菲斯特（Paul Ehrenfest）说，“用五维连续统来取代四维的，然后，为了解释其中一维没有显现的事实，而将它人为地卷缩起来，这是很不正常的事情。”[22]

如果说这些理由还不够，那么物理学家们还有其他理由拒绝这个理论。20世纪30年代，人们认识到世界上除了引力和电磁力外还存在其他力。他们知道强核力和弱核力，因此也该将它们统一进来。但没人知道如何将它们纳入那些统一理论。而且，爱因斯坦还在继续追寻他的统一场论。当时的一些大数学家和物理学家，包括泡利（Wolfgang Pauli）、薛定谔和外尔，也都投身其间。他们发现了其他修正时空几何的方法，从而可以统一引力与电磁力。这些发现依赖于深刻的数学洞察，但仍然没有结果。它们要么没有新的预言，要么预言的现象从来不曾出现过。到了20世纪40年代，爱因斯坦和少数还在追求统一场论的人成了大家的笑柄。

我获得博士学位以后做的第一个工作是1979年在普林斯顿高等研究院。我去那儿的主要原因之一是希望走近爱因斯坦传奇，那时他已经去世24年了。在这一点上，我失望了。除了图书馆的爱因斯坦胸像，他的踪迹已经没有了。这儿没有他的一个学生和追随者，熟悉他的人，只有很少的几个还留在那儿，如理论物理学家戴森（Freeman Dyson）。

我刚到那儿一个星期，戴森（一个很绅士的人）来请我吃午饭。他询问了我的工作，然后问我是否需要他的帮助，在普林斯顿过得更自在。我只有一个问题。我问他，“您能告诉我爱因斯坦到底是个什么样的人吗？”戴森回答，“很抱歉，这个问题我帮不了你。”我很奇怪，接着问，“可是您1947年就来了，直到1955年他去世前一直是他的同事呀。”

戴森解释说，他原本也是怀着认识爱因斯坦的愿望来研究院的。于是，他去见爱因斯坦的秘书杜卡斯（Helen Dukas），请求爱因斯坦见他。会见前一天，他开始担心没有什么特别的问题能与伟人讨论。于是，他从杜卡斯小姐那儿拿回爱因斯坦最近的科学论文，都是关于爱因斯坦构造统一场论的。当晚，戴森读了那些文章，觉得都是些垃圾。

第二天早上，他意识到，尽管不好当面说爱因斯坦的研究是垃圾，但他是有话不得不说的。于是，他失约了。他告诉我，在接下来的8年里，他总是躲着爱因斯坦。

我只能随便问问，“您认为爱因斯坦会替自己辩护，并向你解释他的动机吗？”

“当然，”戴森回答，“但想到那一点时我已经老了。”

除了粒子物理学家们的嘲笑，爱因斯坦和其他几个统一场理论家

们所面临的一个问题是，那种统一太容易了。统一场论不是无处可寻，而是满地鸡毛。它可以通过许多不同的方式来实现，而没有理由说一个比另一个更好。几十年的努力只有一个真正的进步：两种核力被装进来了。结果发现，那只需要添加更多的空间维就行了。当更多的维加入广义相对论时，描述弱核力和强核力的场就显现出来。这很像卡鲁扎的电磁力的情形：必须把额外维的几何固定下来，使它们的几何在时间和空间都不会改变；而且还必须让那些维小得看不见。当所有这些都做好了，将广义相对论方程用于高维空间，就得到必然的方程（即杨一米尔斯方程）。

杨一米尔斯方程隐藏在高维的广义相对论中，这是20世纪50年代才发现的事实，但其意义直到70年代才显现，那时我们已经最终认识了这些方程很好描述了弱核力和强核力。当人们终于发现了那个联系，才有人想复活卡鲁扎一克莱因的思想，但他们没能走得很远。那时候，我们已经知道自然缺少某种对称性——左和右的对称。具体说来，所有中微子都是所谓的左手型的（即它们自旋的方向总是与其线动量的方向相反）。这意味着，如果在镜子里看，将看到一个虚假的世界——其中的中微子都是右手型的。所以，镜像的世界是不可能的世界。但后来发现，在卡鲁扎一克莱因理论描述的世界里，很难解释这种非对称性。

除此之外，高维理论仍然没能提出新的预言。为了得到希望的物理，我们不得不为额外的空间维强加一些条件，这是理论埋下的孽种。实际上，包括的维越多，为了固定它们的几何而付出的代价就越大。空间维越多，自由度也越多——而更多的自由却使那些维的几何偏离我们需要的生成三维世界的力的刚体几何。不稳定性问题变得越来越糟糕了。

而且，只要隐藏维超过1个，就有很多不同的方法来卷曲它们。我们可以用无限多种方式来卷曲多余的维，而不仅仅是将它们化为圆圈。因此，可以有无限多个理论。自然该如何选择它们呢？

在通过额外空间维来统一物理学的努力中，我们一次又一次地遇到相同的情形。有几个解产生了我们看到的世界，但它们只是无限多个可能解的汪洋里的几个不稳定的小岛，其余的地方几乎不可能是我们的世界。一旦为了清除那些解而强加一些条件，就将失去确凿的证据——即统一产生的结果——尽管现在尚未看见，但只要实验家们愿意，是有可能看见的。所以，他们没有可以欢呼的东西，却有很多需要隐藏的东西。

但还有一个更基本的问题，涉及统一理论与量子理论的关系。过去的统一场论的努力发生在量子力学1926年完全建立以前。其实，几个量子理论的支持者早就猜想过额外维与量子理论的有趣联系。但是，到1930年左右，他们发生了分裂。多数物理学家不关心统一问题，而专心于将量子理论用于从材料的性质到恒星的能源过程的大量现象。同时少数坚持统一理论的物理学家则越来越忽略了量子理论。他们（包括爱因斯坦）只管做自己的，似乎普朗克、玻尔、海森伯和薛定谔根本就不存在。他们生活在量子力学革命后的年代，却假装工作在那场革命从未发生的智力世界。在同时的人们看来，他们就像20世纪二三十年代的古怪的俄罗斯流亡的贵族，带着他们精致的旧时尚来到巴黎和纽约，还自以为回到了沙皇时代的圣彼得堡。

当然，爱因斯坦并不仅仅是一个从失去的世界里流亡出来的知识分子。（尽管他曾经真的是从一个失去的世界里流亡出来的知识分子。[23]）他很清楚他忽略了量子理论，但他有自己的理由：他不相信它。尽管他用实在的光子点燃了量子革命的烈焰，却拒绝它的结果。他希望发现一个他能接受的量子现象的更深层理论。他希望他的统一场论能把他引向那个地方。

可是没有。爱因斯坦躲避量子理论的梦破灭了，随他一起消失了。那时，没有多少人关注他，更没有人跟随他。物理学家们认为他们有更好的事情做，用不着玩儿虚幻的统一思想。他们忙着为不断发现的新粒子编目，为新发现的两种基本力琢磨理论。如果谁猜想世界

有更多的小得看不见的空间维，他们会认为他疯了，就像研究UFO一样，是不会有结果的。于是，在理论与实验手拉手前进的这段时期里，没有重大的实验，没有新的预言，也就没有理由注意它。

可是，假如我们明知有那么多障碍却仍然放不下统一场，那么，能用量子理论的语言来建立这些理论吗？回答是一声响亮的“不”。那时候，还没人知道如何协调广义相对论与量子论呢。所有的努力都失败了。当你为空间添加更多的维，或为几何添加更多的扭曲，事情只会更糟而不会更好。维数越多，方程疯得越快，会卷入无限多个量，带来无限多的矛盾。

所以，虽然高维的统一方法很迷人，最终还是因为很好的理由被抛弃了。它没有可以检验的预言。即使这样的理论能出现某个特殊的确实描绘了我们的世界的解，也还有很多没有意义的东西。那很少的几个有意义的解还是不稳定的，很容易演化为奇点或根本不同的世界。最后，它们不能与量子理论协调起来。记住这些理由——因为一个新的统一思想（如弦理论）是成功还是失败，就看它是否能解决这些问题。

我20世纪70年代初开始做物理学的时候，统一引力与其他力的思想已经像连续物质的概念一样破产了。马赫（Ernst Mach）不相信原子，麦克斯韦相信以太，爱因斯坦追寻统一场论。这些都是伟大的思想家们留下的愚蠢的教训。生活就是那么严酷。

## 第四章　统一成为科学

用新的空间维来统一四种基本力的思想失败之后，多数理论物理学家不再相信能将引力与其他力联系起来。这是有道理的，因为引力比其他三种力弱得多。他们的注意力被实验家在粒子加速器中不断发现的基本粒子吸引了。他们要从数据里寻找新原理，希望它至少能统

一不同类型的粒子。

忽略引力意味着人们义退回到了爱因斯坦广义相对论前的时空认识。长远看来，这是很危险的事情，它等于在用废弃的思想来工作。不过这种方法也有一点好处，能把问题大为简化。广义相对论的主要精神在于，空间和时间没有固定的背景几何；忽略这一点意味着我们可以简单选择一个背景。这将我们送回了牛顿的观点，粒子和场发生在空间和时间的固定背景下——背景的性质也是永恒不变的。因此，从忽略引力发展起来的理论是背景相关的。

然而，也不是说一定要重走牛顿的老路。我们可以在爱因斯坦1905年的狭义相对论描述的空间和时间下思考。根据狭义相对论，空间几何是欧几里得几何，即我们在中学学习的那种几何；但根据爱因斯坦的两个基本假定（观察者的相对性与光速的不变性），空间与时间是完全混合的。这个理论容不下引力，但它是麦克斯韦电磁理论的正确框架。

当量子力学完全建立起来时，量子理论家便将注意转向电磁力与量子理论的统一。因为基本的电磁现象是场，所以最后形成的统一理论叫量子场论。因为爱因斯坦的狭义相对论是电磁场的正确框架，所以这些理论也可以看作是量子理论与狭义相对论的统一。

这个问题比将量子理论用于粒子要艰巨得多，因为场在空间每一点都有一个值。如果我们假定空间是连续的——狭义相对论的论断——那么就有无限多个连续变量。在量子理论中，每个变量都满足不确定性原理，其结果是，一个变量测量越准确，它的涨落就越疯狂。无限多个任意涨落的变量很容易失去控制。当我们提出一个理论问题时，千万当心不要有无限多个矛盾的答案。

量子理论家们已经知道，每个电磁波有一个量子粒子，即光子。他们只用了几年时间就弄清了细节，但结果只是自由运动的光子的理

论。下一步还需要容纳带电粒子（如电子和质子），并描述它们与光子的相互作用。这个目标是一个完全和谐的量子电动力学理论（QED），是非常具有挑战性的。QED是日本物理学家朝永振一郎（Sin-Itiro Tomonaga）在第二次世界大战期间首先解决的，但世界其他地方的人到了1948年才知道那个消息。那时，QED已经被年轻的美国物理学家费曼（Richard Feynman）和施温格（Julian Schwinger）独立构造了两次。

一旦明白了QED，人们要做的事情就是把量子场论推广到强弱核力。这是接下来的25年的事情，其关键是发现两个新原理：第一个原理确定电磁力与核力有什么共同的地方，叫规范原理。正如我下面要讲的，它导致了那三种力的统一。第二个原理解释为什么三种统一的力会显得那么不同。它叫自发对称破缺。这两个原理共同形成了粒子物理学标准模型的基石。后来，人们应用它们发现，像质子和中子那样的粒子不是基本粒子，而是夸克组成的。

质子和中子各有三个夸克，而其他那些叫介子的粒子只有两个夸克（更恰当说是一个夸克和一个反夸克）。这是60年代初，加州理工学院的盖尔曼（Murry Gell-Mann）和日内瓦欧洲核子研究中心（CERN）[24]的茨威格（George Zweig）独立发现的。不久，斯坦福直线加速器实验中心（SLAC）的贝约肯（Ja mes Bjorken）和加州理工学院的费曼提出了实验建议，实验后来在SLAC进行，证明了质子和中子确实由三个夸克组成。

夸克的发现是迈向统一的重要一步，因为质子、中子和其他粒子的相互作用非常复杂，而夸克之间的相互作用有可能很简单，质子和中子的外在的复杂性源于它们是复合体。这种观念以前得到过证明：尽管分子间的力很复杂，组成它们的原子之间的力却很容易用电磁学来理解。有了这个思想，理论家们就不打算在基本的层次上去认识质子和中子间的力，而是去探究影响夸克的力。这是还原论在起作用——那是一个古老的策略，认为决定部分的法则通常比决定整体的法

则更简单——结果成功了，发现了在强弱两种核力与电磁力之间存在深层的共性。三个力原来都是简单而强大的规范原理的结果。

规范原理最好是通过物理学家的对称性来理解。简单说，对称是一种不改变事物相对于外在世界的行为的操作。例如，如果你旋转一个球，你不会改变它；它仍然是球。所以，当物理学家谈对称时，指的就是空间里的不改变实验结果的操作（如旋转）。不过，他们也可以谈我们施加在实验上的不改变结果的任何形式的改变。例如，假如我们有两群猫——东边一群，西边一群——来测试它们的弹跳能力。如果猫的平均跳越没有区别，我们就说猫的跳跃在交换东西两群猫的操作下是对称的。

为了更简单、更理想地说明这一点，我们再看另一个例子。考虑一个实验：将一束质子加速，然后瞄准由某些原子核组成的目标。我们来观察质子从核子目标散射后形成的模式。接下来，我们用中子取代质子，但不改变能量或目标。在某些情形，散射的模式几乎不会改变。我们可以说，这个实验揭示了力以相同方式作用于质子和中子。换句话说，用中子替代质子的行为是一种对称——粒子与目标核子之间的相互作用力的对称。

认识对称性是了不起的事情，因为它能告诉我们力的知识。在第一个例子中，我们认识到引力对猫的作用与猫的出身无关；在第二个例子中，核力不能区分质子与中子。有时，我们从对称只能得到力的这样一些部分知识。但也有特殊的时候，对称能完全确定力。所谓规范力就属于这种情形。我不想具体讲它的过程，因为不需要。[25]但我们应该知道，认识对称可以确定一个力的所有性质，这是20世纪物理学最重要的发现之一。这种思想也就是规范原理的精神。[26]

关于规范原理，我们确实需要了解两件事情。一是规范对称的力是通过所谓规范玻色子传递的；二是电磁力和弱力都属于这种类型的力。对应于电磁力的规范玻色子就是光子；把夸克束缚在一起的强力

的规范玻色子叫胶子；而对应于弱力的规范玻色子就不那么好听了——干脆就叫弱玻色子。

规范原理是“美妙的数学思想”，我们在第三章提到过，它是外尔1918年提出的，原是为了统一引力与电磁力，可惜失败了。外尔是对物理学方程考虑最深刻的大数学家之一，正是他认识到麦克斯韦理论的结构完全可以用规范力来解释。50年代，人们怀疑是否其他场论也能用规范原理来构造。结果表明，确实可以在不同基本粒子对称性的基础上构造场论。这些理论现在称作杨—米尔斯理论，是用其创立者的名字命名的。[27]起初，人们不明白这些理论有什么关系。它们描述的新力与电磁力一样，具有无限大的作用范围，物理学家知道两种核力都只在短距离内发生作用，因此似乎不可能用规范理论来描述。

理论物理学之所以既是科学也是艺术，就在于最好的理论家有第六感，能判断哪些结果可以忽略不管。于是，在60年代初，玻尔研究所的博士后格拉肖（Sheldon Glashow）提出，弱作用的确能用规范理论来描述。他只是简单地假定存在某种未知的机制限制了弱力的作用范围。如果力的范围问题解决了，弱力就能与电磁力统一。但仍然面临着一个大问题：我们怎么能统一像电磁力和强弱核力那样表现悬殊的力呢？

我们在这儿看到的是一个困扰着每个统一思想的一般性问题。你想统一的现象五花八门——否则其统一也就一点儿也不稀奇了。所以，即使你找到了它们背后隐藏的统一，也还需要明白它们为什么会有那么不同的表现。

我们前面说过，爱因斯坦用了一种奇妙的方式来解决狭义和广义相对论问题。他认识到现象的区别不是现象的内在特征，而完全是因为需要从小同的观察者的角度来描述现象。电与磁、运动与静止、引力与加速度，都是爱因斯坦以这样的方式统一起来的。因此，观察者感觉的现象之间的区别是偶然的，因为它们只代表了观察者的观点。

60年代，有人对这个一般性问题提出了不同的解决方法：被统一现象之间的区别是偶然的，但并非因为观察者的特殊观点。相反，物理学家起初就有了基本的发现：物理学定律也许具有某种对称性，而定律适用的世界却没有表现与之相关的特征。

我先用我们的社会法律来说明这一点。法律面前，人人平等，我们把这作为法律的一种对称性。用一个人代替另一个人，也不能改变他们要遵从的法律。每个人都要纳税，每个人开车都不能超速。但法律面前的这种平等或对称性，不需要也不会要求我们有相同的环境。有的人比其他人富有，并不是每个人都有小汽车，而有车的人也不是都那么想超速的。

而且，在理想的社会里，人人机会均等。遗憾的是现实社会并非如此。倘若真是那样，我们就可以从初始的机会均等来谈某个对称性。随着我们的成长，那种初始的对称性也离开了。当我们20岁时，机会就各不相同了。有的人会成为钢琴家，还有的人会成为运动员。

我们可以将这种差别归结为初始的对称性随着时间的流逝而破缺了。认为平等是一种对称性的物理学家会说，我们生来具有的对称性被后来的境遇和选择破坏了。在某些情形，很难预言对称会以什么方式破缺。我们知道它一定会破缺，但从幼儿园是看不出来的。在这样的情形下，物理学家说对称是自发破缺的。我们这样说的意思是，对称破缺是必然的，但具体如何破缺是高度偶然的。自发对称破缺是粒子物理学标准模型基础的第二个大原则。

再看一个人类生活的例子。作为老师，我有时参加新生见面会。看着他们从不同的地方走到一起来，我就想，在接下来的几年里，他们有的会成为朋友，有的会成为恋人，还有的可能结成连理。而此刻，他们萍水相逢，满屋子充满着某种对称性；未来的朋友或夫妻都潜伏在这一群人中间。但对称性肯定会被打破，就像在众多的人际关系中必然会生出友谊。这也是自发对称破缺的一个例子。

不论社会的还是自然的，现实的世界都必然会打破各种可能性之间的对称，从而形成世界的结构。这种必然性的特征之一就是对称与稳定之间的平衡。在对称的状态下，所有的人都可能成为朋友或恋人，那是不稳定的。在现实中，我们必须做出选择，从而使状态更为稳定。我们以不稳定的潜在的自由换取稳定的现实的经历。

物理学也是如此。普通的例子是让铅笔竖立在笔尖，这是对称的，即在平衡时，各方向是一样的。但它是不稳定的。铅笔倒下时（肯定会倒下的），它会随机地倒向一方，从而打破对称。一旦铅笔倒了，它就稳定了，但不再表现对称——尽管对称还藏在背后的物理学定律中。这些定律只描述可能发生什么，而决定现实世界的定律还涉及如何从众多可能性中选择一个来实现。

自发对称破缺的机制可以发生在自然的粒子之间的对称性中。当破缺发生在规范原理下产生自然力的那些对称性时，会使那些力表现不同的性质。力就这样区分开了，它们可以有不同的作用范围和强度。对称破缺前，所有四种基本力都和电磁力一样有无穷的作用范围。但对称破缺后，有的力（如两个核力）的范围就变成有限的了。正如前面说的，这是20世纪物理学的最重要的发现之一，因为它和规范原理一起统一了表现迥然不同的基本力。

结合自发对称破缺与规范理论的思想是恩格雷特（Francois Englert）和布劳特（Robert Brout）1962年在布鲁塞尔提出的，几个月后，爱丁堡大学的希格斯（Peter Higgs）又独立发现了它。它本该叫EBH现象，但遗憾的是，人们通常称它为希格斯现象。（科学中的有些事物以最后而不是第一个发现它的人的名字命名，这样的例子还有很多。）他们三位还证明，存在一种粒子，是自发对称破缺的产物。这种粒子叫希格斯玻色子。

几年后，1967年，温伯格（Steven Weinberg）和巴基斯坦物理学家萨拉姆（Abdus Salam）独立发现，可以结合规范原理与自发对

称破缺来构造一个具体的统一电磁力与弱核力的理论。这就是以他们名字命名的理论：弱电力的温伯格-萨拉姆模型。这当然是一个值得欢呼的有具体结果的统一理论；它很快就预言了新现象并成功得到验证。例如，它预言应该存在类似传递电磁力的光子那样的传递弱核力的粒子。那样的粒子有三种，叫W＋，W-和Z。这三种粒子都发现了，而且具有预言的性质。

在基本理论中运用自发对称破缺对后来产生了深远的影响，不仅影响自然律的发现，而且影响我们认识自然律是什么。在此之前，人们认为基本粒子的性质决定于永恒的自然法则。但在自发对称破缺的理论中，出现了一种新的元素，即基本粒子的性质部分依赖于它的环境和历史。这种对称可以通过不同的方式破缺，取决于密度和温度等条件。更一般地说，基本粒子的性质不仅依赖于理论的方程，也依赖于方程的什么解适用于我们的宇宙。

这是物理学与通常的还原论分道扬镳的标志。在还原论看来，基本粒子的性质是永恒的，由绝对的定律确定。而现在看来，基本粒子的很多甚至全部性质，可能都是偶然的，取决于我们如何根据我们在宇宙的位置或我们所处的特殊时代来选择定律的解。不同区域的解可能是不同的，甚至会随时间变化。

在自发对称破缺中，有一个物理量的数值标志着对称的破缺和破缺的方式。那个量通常是一个场，叫希格斯场。温伯格-萨拉姆模型要求希格斯场存在而且表现为新的基本粒子（即所谓的希格斯玻色子），传递与希格斯场相伴随的力。在电磁力与弱力的统一的所有预言中，只有这一点还没得到证实。困难之一在于理论不能准确预言希格斯玻色子的质量；那是理论要求的自由常数之一。人们设计了很多实验来寻找希格斯玻色子，但结果是，假如它存在，其质量必然大于质子质量的140倍。未来加速器实验的主要目标之一就是寻找那样的粒子。

70年代初，规范原理被用到了夸克间的强核力，也发现了与那种力相应的规范场。形成的理论叫量子色动力学（简称QCD）。（用“色”来区别三种不同形式的夸克，是出于好玩儿。）QCD也经历了严格的实验验证，它与温伯格-萨拉姆模型一起构成基本粒子物理学的基础。

三种自然力都是一个统一原理（即规范原理）的不同表现，这个发现是迄今为止的理论粒子物理学的最深刻的成就。完成这个发现的人是真正的科学英雄。标准模型是成百上千的人经过几十年艰辛而痛苦的实验和理论工作的结果。它完成于1973年，30年来经过了众多实验的检验。我们物理学家当然为它感到骄傲。

可接下来的事情就不妙了。现在，三个力被认定为同一个原理的不同表现，显然它们是统一的。然而，为了统一所有粒子，我们需要一个能囊括它们的更大的对称性。然后应用规范原理，生成那三种力。为了区分所有的粒子和力，我们这样来确定对称性：系统的任何组织形态在对称下是不稳定的，而稳定的形态是不对称的。这一点不难做到，因为我们前面讨论过，对称状态本来就是不稳定的。于是，囊括所有粒子的对称性会自发破缺。实现了这一点，三个力才正好表现出我们看到的那些性质。

大统一的思想不仅是要把力统一起来，还要寻找一种对称性将夸克（强力决定的粒子）转化为轻子（弱电力决定的粒子），从而统一两种基本粒子，最后只留下一种粒子和一个规范场。最简单的大统一候选者是SU（5）对称性，名称的意思是5种粒子通过对称性重新组合：三种颜色的夸克（每种夸克都有）和两种轻子（电子和中微子）。SU（5）不但统一了夸克和轻子，而且是无比精妙的统一，精确解释了标准模型的一切，还使许多以前任意出现的东西成为必然的结果。SU（5）解释了标准模型的所有预言，甚至还提出了新的预言。

其中一个新预言是，必然存在从夸克转化为电子和中微子的过程，因为在SU（5）中，夸克、电子和中微子不过是同一种基本粒子的不同表现。我们已经看到，当两种事物统一时，就必然有一种新物理过程将其中一种事物转化为另一种。SU（5）实际上预言了一种类似放射性衰变的过程。这是一个神奇的预言，是大统一的一个特征。它是理论要求的，也是理论独特的地方。

夸克衰变为电子和中微子有着可见的结果。包含夸克的质子不再是质子，它分裂为更简单的东西。于是，质子不再是稳定的粒子——它们会发生某种放射性衰变。当然，假如衰变太频繁，我们的世界也将发生分裂，因为每个稳定的事物都是质子组成的。所以，即使质子要衰变，其衰变率也是非常小的。那也正是理论预言的：大约每1033年才有一个质子衰变。

但是，尽管这种衰变效应很小，却可以做实验来检验，因为世界上有大量质子。所以，在SU（5）中，我们有了最好的一类统一理论，它带来了惊人的结果，而不与我们知道的或可以马上验证的东西矛盾。为了克服质子衰变稀有的困难，我们可以做一个装满超纯水的大池子，这样，池子里的质子可能每年都有几个发生衰变。我们还必须让池子躲避宇宙线，因为宇宙线时刻在轰击地球，能将质子打碎。然后，因为质子衰变产生巨大能量，我们还必须在水池中遍布探测器，等着衰变的发生。资金有了，大水池建在地下深处，我们耐心等着结果。

25年过去了，我们还在等待。没有质子衰变。我们等了很长时间才明白SU（5）大统一是错误的。思想很美，但自然似乎不喜欢它。

最近，我碰到研究生院的朋友爱得华·法尔西（Edward Farhi），现在是麻省理工学院（MIT）理论物理中心主任。我们有20年没有认真谈过了，但我们觉得有很多话要说。我们一直在想，在我们获得博士学位以来的25年里，粒子物理学发生了什么，没发生什么。

爱德华对粒子物理学有过重要贡献，但现在主要从事方兴未艾的量子计算机研究。我问他为什么，他说量子计算不像粒子物理学，我们在那儿知道原理是什么，能认识它们的意义，能做实验来检验我们的预言。他和我都在思考，读研究生时令我们兴奋的粒子物理学，是从什么时候开始沉寂下来的。我们都认为，转折就在于我们发现质子并没有在SU（5）大统一理论预言的时间内衰变。“我本想用自己的生命打赌——哦，也许不是我的生命，你明白我的意思——质子会衰变的”，他说，“SU（5）是个美妙的理论，一切都井然有序——可后来发现它错了。”

其实，我们也不会低估负结果的意义。SU（5）是我们所能想象的最美妙的统一夸克与轻子的方式，它以简单的方式归纳了标准模型的性质。即使25年后，我仍然为SU（5）的失败感到惊讶。

并不是说我们理论家很难避免眼下的失败。只需要给理论添加几个对称性和粒子，就可以出现更多的可以调节的常数。有了这些可调节常数，就可以随意调整质子衰变的速率。这样，我们就可以很容易地避免实验的失败。

如果那样，理论就被破坏了，我们也就不可能看到一个深刻的新思想的惊人而独特的预言。最简单形式的大统一模型预言了质子的衰变速率。如果大统一是对的但更加复杂，能随意调节质子的衰变速率，那么它就不再是解释性的理论了。我们原本希望统一能解释标准模型里的常数值，但SU（5）（如果正确的话）却引进了新的常数，而且，为了避免与实验矛盾的结果，还需要人工调节那些常数。

这是前面讲过的一般性教训的又一个例证。当我们统一不同的粒子和力时，就可能给世界引进了不稳定性。这是因为出现了新的相互作用，统一的粒子要通过它们才能相互转化。这些不稳定性的确是无法避免的；事实上这些过程恰好是统一的证明。唯一的问题在于，我们不知道自己处于什么境地：我们也许运气好——如标准模型的情

形，有明确的很快得到验证的预言；但也许很倒霉，为了隐藏不需要的结果而不得不编造理论。这就是现代统一理论的尴尬。

# 第五章　从统一到超统一

第一代大统一理论的失败给科学带来了至今犹在的危机。20世纪70年代前，理论与实验手拉手前进，新思想在几年或顶多10年内就能得到验证。从18世纪80年代到20世纪70年代，我们关于物理学基础的认识，大概每10年就有一次大的进步。而每一次进步，理论都补充了实验。但自20世纪70年代末以来，我们对基本粒子物理学的认识还没有一个真正的突破。

当一个伟大的思想失败时，总有两种不同的应对方式。我们可以降低标准，先回到从前的知识积累，而不着急用新的理论和实验工具去探索知识的边缘。许多粒子物理学家就是这样做的。结果是标准模型很好地通过了实验验证。过去25年影响最大的发现是中微子具有质量，但这个现象可以通过微调标准模型来满足。除此而外，模型没有任何修正。

对大思想失败的另一种应对方式是找一个更大的思想。开始可能只有几个人走这条路，后来人会越来越多。这是我们不得不走的路线；迄今为止，这些新思想还没有得到实验的支持。

这些年提出和研究过的大思想中，有一个赢得了最多的关注，那就是所谓的超对称。假如它是对的，就可能成为相对论和规范原理那样的我们认识自然的基础。

我们已经看到，这些统一发现了隐藏在原来我们认为不同的各方面之间的联系。空间和时间最初是两个截然不同的概念，狭义相对论统一了它们。几何与引力过去也是毫不相干的，但广义相对论统一了它们。不过仍然存在两大类事物，构成我们生存的世界：构成物质的

粒子（夸克、电子等等）和相互作用的力（或场）。

规范原理统一了三种力，但我们还剩下两样不同的东西：粒子和力。从前有两个努力以统一它们为目标：以太理论和统一场论，但都失败了。超对称是第三个。

量子论告诉我们，粒子是波，波也是粒子，但这并没有统一粒子和力。原因是，量子论还存在两大类基本实体：费米子和玻色子。

构成物质的所有粒子，如电子、质子和中子，都是费米子。所有力都由玻色子组成。光子是玻色子，如W和Z等伴随着其他规范场的粒子，也是玻色子。希格斯粒子也是玻色子。超对称提供了一种统一这两类粒子（费米子与玻色子）的方法。那是一种很新奇的方式，它假定每个已知的粒子都有一个我们尚未看见的超对称伙伴。

大致说来，超对称是一个过程，通过它可以在某些实验中以玻色子代替费米子，而不会改变各种可能结果的几率。这需要很高的技巧，因为费米子与玻色子有着非常不同的性质。费米子要服从不相容原理，那是泡利在1925年提出的，意思是两个费米子不能同时占据相同的量子态。就因为这一点，原子里的电子并不都处在能量最低的轨道。一旦有一个电子处于某轨道（或量子态），就不能在同一个状态放另一个电子。泡利不相容原理解释了原子和材料的很多性质。然而，玻色子的行为却相反：它们喜欢共享一个状态。当我们看到一个光子处于某个量子态时，就可能看到别的光子也在那个态。这种亲和性解释了场（如电磁场）的很多性质。

树立一个理论，能以费米子代替玻色子而得到稳定的世界，乍看起来是很疯狂的想法。但不管怎样，四个俄罗斯人——利希特曼（Evgeny Likhtman）和戈尔方德（Yuri Golfand）在1971年，阿库洛夫（Vladimir Akulov）和沃尔科夫（Dmitri Volkov）在1972年——发现他们可以写出一个具有那种对称性的和谐的理论，那就是我

们现在说的超对称性。

那时，西方科学家与苏联科学家素无往来。苏联科学家难得出国旅行，在非苏联杂志上发表文章也是障碍重重。多数西方物理学家都不看苏联杂志的译本，于是，苏联人的几个发现没有受到西方的注意。超对称性就是其中之一。

就这样，超对称性曾被多次提出。1973年，两个欧洲物理学家魏斯（Julius Wess）和朱米诺（Bruno Zumino）发现了几个例子。他们的工作比苏联人的幸运，得到了关注，并很快有了发展。他们的一个新理论就是电磁力的推广，统一了光子与一种很像中微子的粒子。超对称的另一个发现与弦理论有关，我们后面要更详细地探讨。

超对称能是正确的吗？最初的形式肯定是不对的，它假定每个费米子都有一个相同质量和电荷的玻色子，这意味着必然有一个玻色子具有和电子一样的电荷和质量。这种粒子，假如存在的话，应该称为超电子。但假如真的存在，我们早该在加速器里看到了。

不过，将自发对称破缺的思想用于超对称，这个问题也好解决。结果是直截了当的。超电子获得了很大的质量，于是比电子重得多。调节理论的自由常数——有很多那样的常数——可以使超电子的质量变得任意大。但任何加速器产生的粒子都有质量上限，这就解释了为什么现有的粒子加速器都没有出现过超电子。事实上我们就是这样解释的。

注意这跟我们讲过的其他故事有着相似的地方。在这些故事里，都有人提出新的统一，也导出了重要的实验结果，不幸的是实验不符合理论。于是，科学家将理论复杂化，在其中加入几个可调常数。最后，他们改变常数，隐藏错误的预言现象，从而解释为什么统一（如果正确的话）没有得出任何观察的结果。但这种操作使理论很难证伪，因为我们总可以通过改变常数来解释任何负结果。

从超对称的故事我们看到，它从一开始就是为了隐藏统一的结果。这并不意味着超对称没用，它确实解释了为什么在经过了30多年的发展之后，还没有明确的可以检验的预言。

我只能想象魏斯、朱米诺和阿库洛夫（那几个俄罗斯人中唯一健在的）会有什么感觉。他们也许做出了他们那一代的最重要发现，也许只是发明了一个与自然无关的理论玩具，至于到底是哪种情形，至今还没有证据。在过去的30年里，每个新基本粒子加速器开始运行时，要做的第一件事情就是寻找超对称预言的粒子。一个也没找到。常数不断向上调整，我们也等着下一次实验。

今天，我们都在盯着CERN正在建设中的巨型重子对撞机（LHC）。如果计划进展正常，它将在2007年运行。粒子物理学家们都希望这个机器能帮助我们摆脱危机。最重要的是，我们想LHC能看到希格斯粒子，即携带着希格斯场的大质量玻色子。如果它不能，我们的麻烦就大了。

但风险最大的思想还是超对称。如果LHC发现了超对称，其创立者理所当然会赢得诺贝尔奖。如果没有，就该有纸帽子带了——不是给他们，因为创立新理论是没有什么羞耻的，而是给我们这一代人，因为我们的一生都在扩张那个理论。

LHC承载了太多的希望，因为它的发现能让我们更好地认识第一章提出的五大问题之一：如何解释标准模型的自由常数值？为说明这一点，我们需要理解这些数值的一个突出特征，即它们不是很小就是很大。一个例子是力的强度之间的差别。两个质子之间的电斥力比它们之间的引力强大约38个数量级。粒子质量的悬殊也很大。例如，电子质量是质子的1/1800。如果存在希格斯玻色子，其质量至少是质子的120倍。

为了概括这些数据，我们似乎可以说粒子物理学是有等级的，而

不是民主的。四种力的强度悬殊，从强到弱（即从强核力到引力）形成等级。物理学中的不同质量也形成等级。最顶层的是普朗克质量，它是量子引力效应发生作用时的能量（记住，质量与能量其实是同一个东西）。比普朗克质量大约低4个数量级，是另一个尺度，应该看到电磁力与核力的差别。在那个能量（叫大统一尺度）进行的实验看不到三个力，而只有一个力。比普朗克尺度小16个数量级是TeV尺度（即1012电子伏特），弱核力与电磁力就在这儿统一，因而叫弱相互作用尺度。在这个区域我们应该看到希格斯玻色子，许多理论家还希望看到超对称。LHC就是为了探测这个能量尺度下的物理学而建造的。质子质量比它低3个数量级，再低3个数量级就到电子，而也许还要低6个数量级才到中微子。这样继续小下去，最底层的就是真空能量，即使没有物质，它也存在于整个空间。

这是一幅美妙而疑惑的图像。大自然为什么会那么等级森严？为什么最强与最弱的力悬殊那么大？为什么质子和电子质量比普朗克质量或大统一尺度小那么多？这就是通常所说的等级问题，我们希望LHC能为它带来一点光亮。

那么我们通过LHC能看到什么呢？这是自20世纪70年代初标准模型成功以来粒子物理学的中心问题。理论家盼望LHC已经30年了，我们准备好了吗？令人沮丧的是，没有。

假如准备好了，我们就可以令人信服地预言LHC能看到什么，而我们只需要等着检验就行了。假如我们真的完全认识了粒子物理学，而地球上几千个最聪明的人竟然说不出下一个伟大实验会发现什么，就太奇怪了。但是，除了希望看到希格斯玻色子以外，我们确实提不出什么明确的预言。

你大概以为，既然没有共识，至少总该有几个竞争的理论能提出这样的预言吧。实际情况要坏得多。我们手头真的有几个不同的统一，而且都有一定的成绩，可没有哪一个特别显得更简单、更有解释

能力，也看不出成功的迹象。为什么过了30年我们还不能将理论打扫干净呢，我们需要更仔细地来看等级问题。为什么质量和其他常数有那么大的悬殊？

等级问题面临两个挑战。第一个是什么决定着常数，是什么导致了那么大的差别？第二个是常数是如何固定下来的？这个稳定性问题很令人困惑，因为量子力学有一个奇怪的趋向，要把所有质量拉到一起来，趋近普朗克质量。这里我们不必讨论为什么，但其结果是，我们调节常数的摁钮仿佛由一直绷紧的橡皮筋联系着。

结果，我们可以在标准模型里保留常数的巨大差别，但这要求精确选择常数。我们希望的实际质量的悬殊越大，理论家必须越精细地调节其内禀质量（即没有量子效应时的质量），将它们截然分开。至于如何精细，要看粒子的类型。

规范玻色子的问题不是很大，对称性基本上消除了橡皮筋对其质量的拉扯。不论是否考虑量子效应，光子（携带电磁力场的玻色子）都没有质量，所以它不存在问题。组成物质的粒子，如夸克和轻子，也没有问题。它们来自量子效应的那部分质量，正比于其内禀质量。如果内禀质量小，总质量也小。于是我们说，规范玻色子和费米子的质量是受保护的。

问题出在不受保护的粒子，在粒子物理学的标准模型里，那意味着希格斯粒子，而且只有希格斯粒子。原来，为了防止希格斯质量被拉向普朗克质量，我们必须把标准模型的常数精确调整到小数点后面32位。如果有一位数字不精确，希格斯玻色子最后都会比预言的质量大得多。

于是，挑战落到了希格斯粒子——就是要将它做小。1975年以来探讨过的许多物理学思想都是为着这样一个目的。

驯服希格斯粒子的方法之一是假定它不是基本粒子。如果它是不那么狂野的粒子构成的，问题就解除了。希格斯玻色子由什么组成，人们提出过几个设想。最精致也最贫乏的理论假设希格斯玻色子是很重的夸克或轻子的束缚态。它不添加任何新东西——没有新粒子和需要调节的新参数。这个理论只是假定重粒子以新的方式黏结。这种理论的唯一问题在于很难通过计算验证它、发现新结果。20世纪60年代它刚提出时，超出了我们的实验能力，今天依然如此。

差不多同样精致的另一个假说认为，希格斯玻色子由一种新夸克构成，它不同于组成质子和中子的夸克。起初，这看起来像是解决问题的一种“人工”方案，因而那种夸克被称为“拟夸克”（techniquark）。束缚它们的是一种新力，类似束缚质子和中子的夸克的强核力。在量子色动力学中，力有时叫“色”，于是这种新力被当然地称作“拟色”（Technicolor）。

这个想法容易计算。问题是很难让理论满足观察的各种现象。不过，那也不是完全没有可能，因为它有许多变量。多数变量被排除了，还剩下几个。

第三种假设是将所有基本粒子变成复合粒子。20世纪70年代后期，有几个人在研究这种想法。那是很自然的事情：既然质子和中子由夸克组成，为什么不继续下去呢？也许还有更深层的结构，夸克、电子、中微子，甚至希格斯玻色子和规范玻色子，都是由更基本的粒子组成的，我们可以称它们为“前子”（preon）。这种理论很优美，实验那时已经为我们发现了45种基本费米子，而它们都可以通过两种前子的组合来构成。

而且，这些前子模型解释了观察到的但标准模型没有解释的某些特征。例如，夸克有两种似乎不相干的性质——色与荷。每种夸克表现三种状态（“色”），这个三重态为规范理论提供了需要的对称性。但为什么是三色呢？为什么不是二或四？每个夸克还带有电荷，以电

子电荷的1/3和2/3的形式出现。每种情形都有数字3，意味着色与荷这两种性质可能有共同的起源。不论标准模型还是弦理论（就我所知），都没能说明这种巧合，但前子模型可以非常简单地解释它。

遗憾的是，前子理论也有无法回答的重大问题。那些问题牵扯到什么力把前子粘结成我们看到的粒子。问题就在于，要让那些粒子在保持小质量的同时也保持它们本来的大小。因为前子理论家不能解决这个问题，前子模型到1980年就消亡了。最近我和一些著名物理学家谈话，他们是那以后获得博士学位的，甚至从来没有听说过前子模型。

于是，把希格斯玻色子变成复合粒子的所有努力都不能令人相信。有时，我们理论家似乎山穷水尽了。假如希格斯玻色子是基本粒子，那么该如何把握它的性质呢？

限制粒子自由的一个办法是将它的行为与另一个行为被约束的粒子捆绑在一起。我们知道，规范玻色子和费米子是受保护的，它们的质量不会任意变化。能有将希格斯粒子与质量受保护的粒子系在一起的对称性吗？如果可以那样，那么最后也许能驯服希格斯。我们所知的唯一能做这件事情的对称是超对称，因为超对称联结费米子与玻色子；因此，在超对称理论中应该存在与希格斯粒子为伴的费米子，叫希格斯微子（Higgsino）。（在超对称理论中，约定在费米子的超对称伙伴前冠以字母“s”，而在玻色子的超对称伙伴后缀以“ino”。）希格斯微子是费米子，所以它的质量会受到来自量子效应的质量的保护。这样，超对称告诉我们，两个超伙伴有着相同的质量。因此希格斯粒子的质量也必然受保护。

这个思想很好地解释了为什么希格斯质量比普朗克质量小。前面说过，这个想法很精妙，但其实也很复杂。

首先，理论不可能是部分超对称的。如果一个粒子有超对称伙

伴，那么所有粒子都有。因此，每个夸克有一个叫超夸克的玻色子伙伴。光子的伙伴是一种新的费米子，叫光微子（photino）。于是，相互作用需要调整，当我们在用光微子取代光子的同时也用超夸克取代所有夸克，不同的可能结果发生的几率是不变的。

当然，也有更简单的可能。我们见过的两个粒子就不能是一对伙伴吗？也许光子与中微子会走到一起？或者，希格斯粒子与电子是一对？在已知粒子中发现未知的关系当然是很美妙的，而且令人信服。

遗憾的是，没有一个理论成功假定了两个已知粒子间的超对称。相反，在所有的超对称理论中，粒子的数目至少多一倍。它们只不过假定每个已知粒子伴随着一个超伙伴。不但有超夸克，也有超轻子和光微子。成对的伙伴还有中微子与超中微子，希格斯微子与希格斯玻色子，引力微子与引力子。成双成对的粒子，仿佛满载着一艘粒子的诺亚方舟。纠缠在这个超子与微子的网络里，我们迟早会把巨人看作小丑，把小丑看作巨人，或者别的什么东西。

不管好坏，大自然不是这样的。前面讲过，没有哪个实验产生过超电子的证据。直到今天，似乎也没出现过超夸克、超轻子或超中微子。世界有大量的光子（每个质子对应着十亿多个光子），但没人见过哪怕一个光微子。

问题的解决是假定超对称是自发破缺的。我们在第四章讨论过对称是怎么自发破缺的。这种自发破缺可以推广到超对称。我们可以构造这样的理论：在它描述的世界里，力是超对称的，但那些定律却经过了精心的调节，从而使最低能量状态——即对称性消失的状态——不是超对称的。结果，不需要粒子的超对称伙伴具有相同的质量。

这就产生一个丑陋的理论。为了打破对称，我们必须添加类似希格斯粒子的粒子。它们也需要超伙伴。还有更多的自由常数，可以调节来描述它们的性质。接着，我们不得不调节理论的所有常数，以满

足所有的新粒子都有很大的质量，当然也就看不见了。

对基本粒子物理学的标准模型做这样的事情，不需要添加假设，可以得到一个精巧的结果，叫最小超对称标准模型，简称MSSM。我们在第一章讲过，原来的标准模型大约有20个需要人工调节的自由常数，通过调节它们才能得到与实验一致的预言。MSSM增加了105个常数，为了保证理论与实验一致，理论家可以自由调节它们。假如理论是正确的，那么上帝就成了玩儿杂耍的。他喜欢键盘多的乐器，喜欢16条缆绳的帆船，那才好调整每个帆的形状。

当然，自然也许喜欢这样。理论也可能真的解决常数的调节问题。这样的话，将常数从20个增加到125个，得到的结果是，没有一个新常数需要像原来的常数那样用心调节。尽管如此，有那么多需要调节的常数，实验家很难检验或否定理论。

这些常数有很多安排，对它们来说，超对称都是破缺的，而每个粒子都有不同于其超伙伴的质量。为了隐藏看不见的另一半，我们不得不调节常数，使看不见的粒子质量远远大于我们看见的粒子的质量。我们必须要让这一点正确，因为假如理论预言了超夸克比夸克轻，我们就会有麻烦。不必担心，我们有很多不同的方式来调节常数，以保证我们没有见过的所有粒子都会因为很重而看不见。

如果需要解释这样的常数调节，那么理论必须解释为什么希格斯玻色子具有我们想象的大质量。我们已经说过，即使标准模型也没精确预言希格斯粒子的质量，但它应该比质子重120倍。为了预言这一点，必须调整超对称理论，使超对称性能在这个质量尺度下恢复。这意味着看不见的超伙伴大约都有这个尺度的质量，如果真是如此，LHC应该看到它们。

许多理论家希望LHC将要看到的，是大量可以解释为丢失的超伙伴的粒子。如果LHC真的看到了，那当然是理论物理学家30年来的胜

利。然而我要提醒大家，还没有明确的预言。即使MSSM是正确的，也有很多不同的方法来调节那125个参数以满足我们已知的事实。这至少生出十多种不同的图像，对LHC能看到什么会做出截然不同的预言。

还有更多的麻烦呢。假定LHC生成了新粒子，考虑到超对称理论有那么多不同的形式，那么即使超对称理论错了，也仍然可能经过调节而满足LHC的第一批发现。为了证明超对称，还需要更多的东西。我们需要发现更多的新粒子并解释它们。而它们也许并不都是已知粒子的超伙伴。一个新粒子可能是另一个尚未发现的新粒子的超伙伴。

证明超对称正确的唯一无懈可击的方法是证明确实存在某种对称性——就是说，对各种可能的实验结果，我们有可能用一个粒子来替代其超伙伴，而结果不发生改变。但这对LHC来说，至少在开始的时候很难实现。所以，即使在最好的情形，我们也需要再等很多年，才能知道超对称是否是常数调节问题的正确解释。

同时，许多理论家似乎都相信超对称。有几点不错的理由认为它是旧的统一思想的进步。首先，希格斯玻色子，假如不是点粒子，似乎不会很大。这就有利于超对称而排除了某些（尽管不是所有）拟色理论。另一点理由来自大统一思想。我们在前面讨论过，在大统一能量尺度下进行的实验不能区分电磁力与核力。标准模型预言存在这样的统一尺度，但需要小小的调整。超对称形式的标准模型带来了更直接的统一图景。

超对称当然是很迷人的理论思想。力与物质的统一思想为基础物理学中最深层的对偶性提供了解决方法。难怪那么多理论家觉得简直难以想象一个不是超对称的世界。

同时，确实也有物理学家担心超对称（如果真的有）早就应该在实验中看到了。在最近一篇论文的引言里，我们看到这样的典型说

法："LEPII[CERN的巨型正负电子加速器]没有发现任何超粒子或希格斯粒子，这个事实引出了另一个问题。"[28]北卡罗莱纳大学著名理论家弗拉姆普顿（Paul Frampton）最近写信给我说，我在过去十多年的一般观察是，多数研究TeV尺度超对称破缺现象的人（有几个例外）都认为，TeV尺度超对称在实验中显现的可能性远小于50%，大概也许只有50%。[29]

不管怎么说，我个人的猜想是，超对称（至少对迄今研究过的形式说）不能解释LHC观测的东西。在任何情形，超对称都由实验决定，不论多么偏爱美学标准，我们都盼着有一个答案能告诉我们它是否是正确的自然图像。

但是，即使发现了超对称，它本身也解决不了我在第一章列举的那五个大问题。它不能解释标准模型的常数，因为MSSM有更多的常数。它也不能选择引力的量子理论，因为主要的理论都与超对称的世界相容。也许暗物质是超伙伴构成的，但我们需要直接检验。

这个更大缺陷的原因在于，虽然超对称理论有了更多的对称性，却没变得更简单。其实它们比对称性少的理论要复杂得多。自由常数的个数没有减少——反倒大大地增加了。它们不能统一我们已经知道的任意两样事物。假如超对称性能揭示两个已知事物背后的共性，当然应该是很迷人的——就像麦克斯韦的电磁统一那么迷人。假如能证明光子和电子，甚至中微子与希格斯微子是一对超伙伴，那就太美妙了。

但任何超对称理论都不是这样的。相反，它们假定了一组新的粒子，使每个粒子与某个已知或未知的粒子为对称伙伴。这种理论成功太容易了。创立一个全新的未知世界，然后建立一个有很多参数（可以调节参数隐藏新的粒子）的理论，是不那么动人的，即使它在技术上引人入胜。做这种理论是不会失败的，因为与现有数据的任何矛盾都可以通过调节常数而消除。只有当它面对实验时，才可能失败。

当然，这并不是说超对称不对。它可能是正确的，如果真是那样，LHC就有可能在未来的几年发现它。但超对称没有我们希望的那些行为，这意味着它的支持者们可能远离了经验科学的大树，而摇摇欲坠地坐在一个小枝丫上。也许正如爱因斯坦说的，哪儿的木头薄就在哪儿打钻，不过那是要付出代价的。

## 第六章　量子引力：岔路

粒子物理学家忽略引力时，几个勇敢者从20世纪30年代就开始思考把引力与飞速发展的量子理论融合起来。在半个多世纪里，做量子引力的先驱屈指可数，也很少有人关注他们。但量子引力的问题不会永久被冷落。在我提出的那五个大问题中，它是真正不容回避的。它不像别的问题，它在寻求一种书写自然律的语言。解决任何其他问题而不先解决它，就像跟一个没有法律的国家进行谈判。

量子引力是真正的追求，思想的先驱者就像寻找新世界的探险家。现在，探险的人多了，有些景观已经清楚地画出来了。人们还发现有的行迹只能通向死地。有的地方在发出光亮，有的地方开始拥挤，这个时候我们还不能说问题解决了。

本书大部分内容写于2005年，正好是爱因斯坦第一个伟大成就发表100周年。这一年有很多纪念活动。对任何人来说，这都是关注物理学的一个很好理由，当然这没有一点儿讽刺的意味。爱因斯坦的有些发现很激进，直到今天也还没得到某些理论物理学家的足够理解，其中最主要的一点就是他的广义相对论关于空间和时间的认识。

广义相对论的主要教诲是，空间几何不是固定不变的。它动态地演化着，当周围物质运动时，它也随时间而变化。空间几何里甚至还穿越着引力波。在爱因斯坦之前，我们在中学学习的欧几里得几何一直被作为永恒的法则：三角形三个内角之和等于180度，这总是正确

的，而且将永远正确。但在广义相对论中，三角形的内角和可以是任意的数值，因为空间几何可以发生弯曲。

这并不意味着存在另一个固定的刻画空间的几何，例如说空间像球或马鞍而不是平面。重要的是几何可以任意改变，因为它在物质和力的影响下随时间演化。我们的定律不是陈述几何是什么，而是决定几何如何变化——就像牛顿定律不是告诉我们物体在哪儿，而是通过确定力对运动的影响来说明它们如何运动。

爱因斯坦之前，几何被认为是定律的一部分。爱因斯坦揭示了空间几何遵从更深层的定律而在时间中演化。

完全理解这一点是非常重要的。空间几何不是自然律的一部分。于是，在那些定律中，没有什么决定空间几何的东西。因此，在解爱因斯坦广义相对论方程之前，我们对空间几何没有任何概念。只有在解方程之后，我们才知道那几何是什么。

这意味着自然律的表达形式不能再假定空间有任何固定的几何。这是爱因斯坦思想的核心。我们前面将其概括为一个原理，即背景独立性。这个原理说，我们可以完全确定自然定律，而不需要对空间几何做任何先验的假定。在过去那种几何固定的图景中，几何可以认为是背景的一部分，是大自然自我表现的不变的大舞台。我们说物理学定律是背景独立的，意思是空间几何不是固定的，而是演化的。空间和时间不是为事物的演化搭舞台，而是从定律中产生出来。

背景独立的另一方面是说不存在地位特殊的时间。广义相对论最根本的就是以事件和事件之间的关系来描述世界的历史。主要的关系涉及因果性。一个事件可以处于一个因果链，引发别的事件。从这个观点看，空间是派生的概念。空间概念其实完全依赖于时间的概念。拿一个时钟，我们可以考虑正午钟声响起时发生的所有事件，它们构成空间。

广义相对论的一个重要观点是不存在特别的计时方式。任何种类的时钟都可以用，只要它说明原因在结果之前。但是因为空间定义依赖于时间，空间的不同定义与时间一样多。我刚才说了空间几何在时间中演化，那不单是对一个普适的时间概念，也是对所有可能的时间概念。这些概念如何作用，是爱因斯坦广义相对论复杂和美妙的一部分。对我们来说，只需要记住一点，即广义相对论的方程告诉我们的是空间几何如何在时间——不是某个时间，而是所有可能的时间——中演化。

实际上，背景独立性的意义不止于此。自然律还有一些方面固定在通常的物理学定律的表述中。但那也许是不应该的。例如，空间只有三维，这个事实就是背景的一部分。是否存在某个更深层的定律不需要我们先验地假定空间的维数呢？在那样的理论中，三维也许作为某个动力学定律的解的结果而出现的。在那样的理论中，甚至空间的维数也可能随时间而改变。假如我们能构造这样的理论，它也许能解释为什么我们的宇宙是三维的。这将是一个进步，因为原来只能假定的东西终于得到解释了。

所以，背景独立的思想在最广泛的意义上是关于如何做物理的一种智慧：构造更好的理论，在这个理论中，现在假定的东西将在某个新定律下演化，从而得到解释。爱因斯坦的广义相对论就对空间几何实现了这个思想。

因此，引力的量子理论的关键问题是：我们能将空间没用固定几何的思想推广到量子理论吗？就是说，我们能使量子理论成为背景独立的吗（至少对空间几何而言）？假如能做到这一点，我们将自动融合引力与量子论，因为我们已经把引力作为动力学的时空几何的一个方面。

于是，有两种方法融合引力与量子论：实现或不实现背景独立性。量子引力领域从1930年起就分化为这两条路线，尽管今天研究的

多数方法都是背景独立的。一个例外是当今多数物理学家走的路线——弦理论。

20世纪最伟大的科学家的最高成就却被多数叫嚷着的追随者们忘到了脑后，这种状况的发生，也许是科学史上最奇异的事情。但我们必须在这儿讲这个故事，因为它对我在绪言里提出的问题是至关重要的。其实，你大概也感到奇怪，既然爱因斯坦的广义相对论得到了普遍认同，为什么还有人想脱离它的核心原理而另创什么新理论呢？答案也在于一个故事，和本书讲的其他许多故事一样，也从爱因斯坦开始。

1916年，爱因斯坦已经认识到引力波存在并携带能量。他立刻注意到，为了与原子物理学相容，引力波携带的能量也应该用量子理论来描述。在关于引力波的第一篇论文里，爱因斯坦说，“看起来，量子理论不仅会修正麦克斯韦的电磁理论，也必将修正新的引力理论。”[30]

不过，虽然爱因斯坦第一个提出了量子引力问题，他最深刻的见解却被那时以来的多数研究者忽略了。怎么会这样呢？

真有一个理由。那时候，还没人知道如何将正在发展的量子理论用于广义相对论。相反，通过间接的路线倒是有可能取得进展。想把量子力学用于广义相对论的人面临着两个挑战。除了背景独立性，他们还必须把握广义相对论是场论的事实。就是说，空间几何有无限多种可能，从而有无限多个变量。

我在第四章讨论过，当量子力学刚完全建立，物理学家就将它用于场论，如电磁场。它们是建立在固定时空背景的理论，因此没有引出背景独立性问题。但物理学家从中学会了把握无限多个变量的问题。

量子场论的第一个成功是QED，是麦克斯韦电磁论与量子论的统一。值得注意的是，1929年，量子力学的两个创立者，海森伯和泡利，在他们的第一篇QED论文里，已经在考虑将他们的工作推广到量子引力。他们显然觉得问题不是太难，因为他们写道，“引力场的量子化——从物理学的角度看是必然的——可以用完全类似于这里的形式来实现，而不会遇到任何新的困难。”[31]

75年过后，我们对那样两个杰出人物如此低估量子引力问题的困难，只能感到惊讶。他们在想什么呢？哦，我知道了，因为自那时以来很多人都有同样的思想，还被它彻底引向了死路。

海森伯和泡利想的是，当引力波很微弱时，可以看作在固定几何下的扰动的微澜。假如在一个宁静的早晨向水池扔一颗石子，它会在平静的水面的固定背景上激起微波。但是，如果在风雨大作的日子看海湾的汹涌波涛，就不能再把它当作对某种固定东西的扰动了。

广义相对论预言宇宙中存在那样的区域，其时空几何在狂乱地变化着，就像拍岸的惊涛。但海森伯和泡利认为，先考虑引力波极其微弱的情形会简单一些，可以将波看作固定背景里的小波动。这样，就可以运用他们为研究电磁场（固定背景下运动的场）发展的方法。将量子力学用于微弱的自由运动的引力场确实不难。结果，每个引力波都能以量子力学的观点视为一个粒子，叫引力子——类似于作为电磁场量子的光子。但是接下来他们遇到了大问题，因为引力波会相互作用。它们可以与任何具有能量的事物发生作用，而它们本身也携带着能量。电磁波没有这样的问题，因为尽管光子与电荷和磁荷作用，它们本身不带荷，因而没有相互作用。两种波之间的这一重要差别正是海森伯和泡利忽略了的东西。

如何描述引力子的自相互作用，一直是一个难以攻克的难题。我们现在知道，失败的原因是没有认真考虑爱因斯坦的背景独立性原理。一旦引力波相互作用，就不能再认为它们是在固定背景里运动。

相反，它们在运行中改变了背景。

20世纪30年代就有几个人明白了这一点。俄罗斯物理学家布隆斯坦（Matvei Petrovich Bronstein）1935年的学位论文大概是关于量子引力的第一篇博士论文。还记得他的人都认为他是当时前苏联最有才华的两个物理学家之一。他在1936年的一篇论文里写道，“为了消除这个逻辑矛盾，必须屏弃我们寻常的时间和空间概念，以更深层的概念来修正它。”接着，他引用了一句德国俗语：“让怀疑它的人付出代价吧。”[32]布隆斯坦的观点得到了同样才华横溢的法国青年物理学家所罗门（Jacques Solomen）的拥护。

今天，几乎每个认真思考量子引力问题的人都会赞同布隆斯坦，可已经过了漫长的70年。原因之一是，即使像布隆斯坦和所罗门那样聪明的头脑也没能逃脱时代的疯狂和愚昧的厄运。布隆斯坦在写了那篇论文一年后，就被前苏联内务人民委员会（NKV D）逮捕了，1938年2月18日被处死。所罗门是法国抵抗组织的成员，1942年5月23日被德国人杀害。他们的思想被历史遗忘了。我一直在研究量子引力，但写完本书才了解到他们。

布隆斯坦的工作被遗忘了，多数物理学家回头去研究量子场论。我在第四章说过，QED直到20世纪40年代末才发展起来。这一成功激发了一些人重新担负起统一引力和量子论的挑战。于是立刻形成两个对立的阵营。一个随布隆斯坦，重视广义相对论的背景独立性；另一个忽略背景独立性，走海森伯和泡利的路线，将量子论用于在固定背景下运动的引力波。

因为背景独立性是广义相对论的一个原理，将它融人与量子论的统一应该是合理的。但结果表明，事情并不那么简单。有些人——如英国物理学家狄拉克（P.A.M.Dirac）和德国人贝格曼（Peter Bergman，曾在普林斯顿做爱因斯坦的助手，从而开始了学术生涯）——确实尝试过构造一个背景独立的量子引力理论。他们发现那真是

一项艰巨的任务。直到80年代中期，这些努力才有了一点成果。不过那时以来，以背景独立观点来认识量子引力已经有了很多进步。多数量子引力的理论家在以不同的背景独立方法进行研究。这些事情我们后面还要讲，因为它们是弦理论的最重要替代者。

但是，当人们在50年代开始走上量子引力路线时，这些前景还没显露一丝迹象。与QED的大踏步前进比起来，背景独立方法取得的有限进展就显得微不足道了。于是，从80年代末以前，多数人都选择了别的路线，试图将QED方法用于广义相对论。这大概是可以理解的。QED建立之后，人们对背景相关的量子理论有了很多认识，但没人知道背景独立的量子理论（假如有的话）像什么样子。

这是引向弦理论的路线，所以我们需要认真追溯。因为理论在30年代就被人遗忘了，所以还要重新发现。后来，引力子理论在德维特（Bryce DeWitt）的博士论文里重新出现了，40年代他是施温格在哈佛的学生。因为德维特的这个成果和许多其他发现，我们认为他是量子引力理论的奠基者之一。

可我们前面讲过，引力子理论还不够。如果引力子仅仅在空间运动，引力子理论是很好的。但如果引力子真的只是那样，就没有引力了，当然也就没有动力学的或弯曲的几何。所以，这不是广义相对论或引力与量子理论的统一。50年代初，当人们又开始研究它们的相互作用时，引力子理论的问题再次出现。从此以后直到80年代，人们为了消除与量子理论原理的矛盾，在引力子的自作用问题上做了大量工作。结果没有一个成功的。

也许我们应该停下来，想想这在我们的生活中会意味着什么。我们谈了30多年的艰苦劳作，进行了大量复杂的计算。想象一下，假如你整天计算自己的个人所得税，算了一个星期，还不能得到正确的结果，那一定是哪儿出了问题，只是你没发现它。假如你这样过了一个月，你还会让它延续一年吗？那么，假如过了20年呢？再假如周围好

几十个人，有朋友，也有对手，都这样过日子，情况会怎样？他们都有自己的生活策略，每个策略到头来都失败了。但是，尝试一个略微不同的方法，或者将两种方法结合起来，也许你能成功。你每年参加一两个国际会议，你把新策略告诉其他疑惑的人。这就是1984年前的量子引力领域的状况。

费曼是第一个攻击这种引力子理论的人。当然应该攻击。他在QED做过那么好的工作，为什么不把同样的方法用于量子引力呢？60年代初，他离开了粒子物理学，花了几个月的时间来看自己是否能把引力量子化。为了让大家真切感受量子引力那时是怎样的一潭死水，我们来看费曼1962年给妻子的一封信，谈的是在华沙举行的一次会议，他在会上还做了报告：

我没从会上得到任何东西，什么也没学到。因为这个领域没有实验，一点儿也不活跃，几乎没有最优秀人物做那些事情。结果来了一大群笨蛋……这对我的血压没好处。记得提醒我，以后不要参加任何引力会议！[33]

不过，费曼还是取得了很好的进展，澄清了与概率（0和l之间的数）有关的一个技术问题。任何肯定发生的事件，我们说它有概率1。所以，任何所有事件发生的概率为l。在费曼之前，没人能使量子引力的各种事件发生的概率总和为1。实际上，费曼只是在1阶近似下计算了概率之和；几年后，德维特指出了如何将其推广到任意阶的情形。大约1年后，两个俄罗斯人，法德耶夫（Ludwig Dmitrievich Faddeev）和波波夫（Victor Nicolaevich Popov）也发现了这一点。他们不可能知道德维特的工作，因为杂志把他的文章寄给了一个专家评审，而审稿人用了一年多的时间才看完。就这样，人们在一点点地解决问题——但即使概率总和等于1，引力子理论在总体上还是不能运行。

这项工作带来一些副产品。同样的方法可以用于标准模型赖以建

立的杨-米尔斯理论。于是，当温伯格和萨拉姆用这些理论来统一弱力和电磁力时，计算的技术已经具备了。结果比量子引力的好。荷兰理论家特胡夫特（Cerardt　Hooft）最终在1971年证明，杨-米尔斯理论作为量子理论是恰到好处的。实际上，特胡夫特和以前的人一样，是为量子引力热身而研究杨-米尔斯理论的。所以，30年的量子引力研究也不完全是浪费，它至少能使我们更聪明地做粒子物理学。

但没有谁能救量子引力。人们尝试过所有的近似方法。因为粒子物理学的标准模型是有效的，许多方法就是为了探究其不同特征才发展起来的。这些方法一个个都用到量子引力问题，但每一个都失败了。不管怎样组织引力波的量子理论，只要考虑了它们相互作用的事实，就会产生无穷大量。不管怎么处理，那些无穷大都不会消除。尽管又经过了多年，发表了更多的文章，出现了更多的博士，举行了更多的学术会议，情况依然如故。到1974年，人们终于明白了，用背景相关方法来结合广义相对论与量子理论是没有意义的。

然而，背景相关方法确实可以做一件事情。原先是为了认识量子理论对引力波的影响而将引力量子化，现在我们可以把问题反过来，看引力对量子现象会有什么影响。为此，我们可以研究量子粒子在引力作用下的时空（如黑洞或膨胀的宇宙）中的运动。自20世纪60年代以来，这方面取得了很多进展。这是一个重要方向，因为有的发现引出了后来的方法（如弦理论）需要解决的疑难。

第一个成功是预言了在引力场迅速变化时会生成基本粒子。这个思想可以用于迅速膨胀的早期宇宙，引出了我们今天熟悉的关于早期宇宙的预言。

这些计算的成功激发了个别物理学家去尝试研究更难的问题，即黑洞对量子粒子和场的影响。这儿的问题在于，虽然黑洞占据着一个急剧演化的时空区域，那个区域却隐藏在视界后面。所谓视界是光线静止的一个叶面，它标志着在那个界线内部的所有光线都会被拉向黑

洞的中心。于是，没有光线能从视界背后跑出来。从视界外面看，黑洞是静止的，但一旦进入视界，那个区域的事物就被拉向越来越强大的引力场。它们终结于一个奇点，那里的事物都是无穷，那里的时间停止了。

联系量子论与黑洞的第一个关键结果是贝肯斯坦（Jacob Bekenstein）在1973年发现的，这个以色列小伙子是惠勒（John Archibald Wheeler）在普林斯顿的研究生。他惊讶地发现黑洞有熵。熵是无序的度量，有个著名的定律（热力学第二定律）说，封闭系统的熵永远不会减小。贝肯斯坦担心的是，假如他拿一个充满热气体的盒子——它会有很高的熵，因为气体分子的运动是随机和无序的——将它扔进黑洞，则气体再也回不来了，于是宇宙的熵将减小。为了挽救第二定律，贝肯斯坦假定，黑洞本身应该有熵，当气体盒子扔进去时，它的熵会增大，这样整个宇宙的熵才不会减小。他通过几个简单的例子就发现了黑洞的熵必然正比于包围它的视界的面积。

这引出一个问题。熵是随机性的度量，而随机运动是热的。难道说黑洞也有温度？一年后，1974年，霍金才证明黑洞实际上肯定是有温度的。他还确立了黑洞视界面积与熵之间的精确比例关系。

霍金还预言了黑洞温度的另一方面特征，对我们后面的讨论很重要，即黑洞温度与其质量成反比。这意味着黑洞行为和普通物体不同。对大多数事物来说，为了将其加热，需要输入能量，“为火焰添加燃料”。黑洞行为正好相反，给它添加能量或质量，可以使它更大，但也使它变冷了。[34]

从此，引力的量子论的每一次努力都要面对这个奇妙的挑战：如何根据第一原理解释黑洞的熵和温度？贝肯斯坦和霍金把黑洞看作量子粒子在其中运动的经典的固定背景，然后基于与已知定律的一致性而进行论证。他们没有把黑洞描述成一个量子力学体系，因为那只能在时空的量子理论下才能做到。因此，任何引力的量子理论的问题就

是要让我们对贝肯斯坦熵和霍金温度有更深刻的认识。

第二年，霍金发现那些结论还有一点疑惑。因为黑洞有温度就会像任何热体一样产生辐射。但辐射会从黑洞带走能量。只要有足够的时间，黑洞的所有物质都将转化为辐射。当黑洞失去能量时，质量会减小。而根据刚才描述的性质，它失去质量会加热，所以黑洞会辐射得越来越快。到过程的最后，黑洞收缩到普朗克质量，因而需要引力的量子理论来预言它的最终命运。

但不管最终命运如何，关于信息的命运似乎都是一个难题。在黑洞的一生中，会吸收大量的携带着大量内在信息的物质。可最后留下的只不过是一个小黑洞和大量的随机而不带任何信息的热辐射。难道信息就这样消失了吗？

这是一个量子引力问题，因为量子力学中有个定律说信息不会丢失。世界的量子描述被认为是精确的，这隐含着一个结论：当所有细节都考虑时，没有信息可以丢失。霍金做了很强的论证，说明蒸发的黑洞会丢失信息。这看来与量子理论矛盾，所以他说自己的论证是黑洞信息悖论。任何可能的引力的量子理论都需要解决它。

20世纪70年代的这些发现是通向引力的量子理论的里程碑。从那时起，我们要度量一个量子引力方法的成功，就看它在多大程度上回答了黑洞熵、温度和信息丢失等疑难问题。

大约这时候，人们终于提出一个有用（至少用过一时）的量子引力思想。它涉及将超对称的思想用于引力，结果就是超引力。

我听过这个新理论的一次早期讲座。那是1975年在辛辛那提召开的一个广义相对论的会议。那时我还只是汉普郡学院的大学生，但还是去了，就想看看人们在想些什么。我还记得芝加哥大学格罗赫（Robert Geroch）的几个精彩演讲，他当时是这个领域的一颗明

星，专门研究无限空间的数学。他那美妙的论证赢得了满堂喝彩。接着，会议结束时，一个叫彼德·范·纽文惠增（Peter van Nieuwenhuizen）的年轻博士后做了演讲。他开始就说要介绍一个崭新的引力理论，吸引了我的全部注意力。

彼德说他的新理论以超对称（当时是统一玻色子与费米子的一种新思想）为基础。从引力波的量子化我们得到叫引力子的粒子，它们就是一种玻色子。但是对具有超对称的系统来说，它必须既是玻色子也是费米子。广义相对论没有费米子，新的费米子必须假定为引力子的超伙伴。“超引力子”不好听，所以改称它们为“引力微子”。

彼德说，由于从未见过引力微子，我们可以自由设计它所满足的定律。为了让理论在超对称下对称，在以引力子代替引力微子时，力不能改变。这就给定律强加了很多约束，寻求那些约束的解需要很艰辛的计算。有两个研究小组差不多同时完成了那些计算。彼德是其中一个小组的成员；另一组中有我未来在哈佛的导师德赛（Stanley Deser），他当时正和超对称的创立者之一朱米诺一起工作。

彼德还讲了这个理论的另一种思想方法。首先考虑空间和时间的对称。普通空间没有特殊的方向，因此其性质在旋转下保持不变。如果我们从一个地方移动到另一个地方，它们也不会改变，因为空间几何是均匀的。于是，空间具有平移和旋转对称性。回想一下我在第四章解释过的规范原理，即在某些条件下对称性可以指示力所服从的定律。我们可以将这一原理用于空间和时间对称性。结果正是爱因斯坦的广义相对论。爱因斯坦当然不是这样建立他的理论的，但如果他还活着，很可能会以这种方式建立广义相对论。

彼德解释说，超对称性可以看作空间对称性的深化。这源于一个深刻而美妙的性质：假如把所有费米子转化为玻色子然后又转回来，那么世界还是从前的样子，只是每样事物都在空间移动了一点距离。我不能在这儿解释为什么会这样，但它告诉我们超对称性与空间几何

有着某种根本的联系。由此，当我们将规范原理用于超对称时，就会得到一个引力理论——超引力。从这个角度看，超引力是广义相对论的深化。

我是这个领域的新人，偶然走进了会场。我没有一个认识的人，也不知道彼德的听众对他要说的有什么想法，但我还是留下了深刻的印象。我回家想，那个小伙子如此朝气勃发，实在太好了。如果他说的是对的，那可真的很重要。

在研究生院的第一年，我选了德赛的课，听他讲新的超引力理论。我发生了兴趣，开始思考，但也感到困惑。它到底是什么意思？它想告诉我们什么？我在那儿结识了一个新朋友，我的同班同学马丁·罗瑟克（Martin Rocek），他也同样感到兴奋。他很快就跟当时在石溪的彼德取得了联系，并且开始与他和他的学生合作。石溪不远，马丁带我去过一次。故事全面展开了，他想给我一个机会，在其中占个好位置。

那就像得到微软或谷歌的职位。罗瑟克、彼德和我遇到的很多人都在超对称和超引力中有过出色的表现。我相信，在他们看来，我的表现就像一个傻子，放过了好机会。

对我（我相信对别人也一样）来说，超对称性与空间和时间理论的融合产生了深刻的问题。我原来从爱因斯坦的著作里学习过广义相对论，如果说我理解了什么，那就是将引力与空间和时间的几何结合起来的方法。那个思想成了我的主心骨。这时我听说自然的另一面也把空间与时间统一起来了——那就是存在费米子与玻色子。我朋友这样说，方程也这样说。可不论朋友还是方程都没告诉我它有什么意思。我不懂那个思想，不明白它的概念。作为这个统一的结果，我也许应该更深入地理解空间、时间和引力，理解费米子和玻色子意味着什么。那不应只是数学——我的自然概念需要改变了。

但事情不是这样的。当我和彼德的学生们在一起时，才发现他们是一群技术娴熟的年轻人，着迷似地做着计算，夜以继日而不知疲惫。他们做的就是构造不同形式的超引力。每一种形式都有比从前更多的对称性，统一着更多的粒子家族。他们在走向一个终极的理论，将把所有粒子与时间和空间统一起来。这个理论只有一个学名：N＝8理论，这儿N是混合费米子与玻色子的不同方式的数量。第一个这样的理论——也就是彼德和德赛向我介绍的那个——是最简单的，N＝1。欧洲有人做了N＝2。我在石溪的一个星期里，那儿的人正向N＝4前进，还想达到N＝8。

他们日夜不停废寝忘食地工作，忍受着工作的单调和沉闷，确信他们正在了解某些将给世界带来变化的新生事物。有人告诉我，他正在加紧工作，因为他相信，如果人人都知道构造新理论原来是那么容易，他的领地就将泛滥了。其实，假如我没记错，那群人真的得到了N＝4，但在N＝8遇到了陷阱。

他们做的事情在我看来一点儿都不容易。计算枯燥乏味，令人思想麻木。而且，计算还必须完全精确：假如某处丢失一个因子2，那么几个星期的气力都白费了。计算的每一行都有几十项，为了能把一行行计算写下来，他们需要越来越大的本子，很快就用所能找到的最大的绘画本来写了。每页纸都密密麻麻写满了小字，每个本子代表几个月的工作。这令我想起“苦行僧”。我害怕了，呆了一个星期就溜走了。

在后来的几十年里，我与彼德、马丁等人的关系相当紧张。他们大概把我当成了失败者，辜负了他们给我的与他们一起共创超引力的机会。如果我加入他们，也许会成为弦理论的领头人之一。而我做的却是走自己的方向，最终使我发现了量子引力的不同的路线。那让事情变得更糟：我不是放弃真理的失败者，而是要成为一个对手的失败者。

当我回顾30年来我所熟悉的那些人们的科学生涯时，越发感到科学生涯的抉择依赖于人的个性。有些人乐于跨越下一步，把一切都献给它，从而为飞速发展的领域做出重要贡献。另一些人可没那么急躁。有些人容易犯糊涂，所以做什么都要反复思量，这要费很长的时间。你大概以为我们比这些人高明，可别忘了爱因斯坦也是属于他们的。根据我的经验，真正令人震撼的新思想方法往往来自这样的人群。还有一些人——我属于这第三类——只顾走自己的路，他们特立独行，只是因为不愿意像有的人那样为了站在赢家的一边而加入某个领域。所以，当我与别人的作为相左时，也不再感到烦恼，因为我发现一个人的性情几乎完全决定了他做什么样的科学。幸运的是，科学需要来自不同类型的人物的霓献。我逐渐认识到，那些能把科学做好的人是因为他们选择了适合自己的问题。

不管怎么说，我脱离了石溪的超引力小组，但没有失去对超引力的兴趣。相反，我比过去更有兴趣了。我相信他们认识了某些东西，但他们走的路线不是我要走的。我理解爱因斯坦的广义相对论，这意味着我知道怎样以简洁的一页或更少的文字来说明它的每一个基本性质。在我看来，如果你真的理解一个理论，就用不着花几个星期的时间来为检验其基本性质而计算。

我和另一个研究生结伴——他是我在汉普郡学院的朋友约翰·德尔（John Dell），在马里兰大学工作。我们想更深入地了解超对称如何成为时空几何的一部分。他找到了几篇数学家科斯坦特（Bertram Kostant）写的关于一种新几何的文章，它们推广了爱因斯坦的方法，加入了一些看起来有点像费米子的新性质。我们在那种新背景下写出广义相对论的方程，超引力的方程就冒出来了。我们就这样有了第一篇科学论文。

大约同时，别人也发展了不同的超引力几何的方法，叫超几何。那时我觉得他们的工具比我们的笨拙（现在也这么看）。它非常复杂，但对某些问题很有成效。它在一定程度上简化了计算，那当然是

令人欣喜的。所以超几何流行了，而我们的工作被遗忘了。约翰和我并不在意，因为两种方法都没能给出我们在找的东西。虽然数学是成功的，但没有带来任何概念的进步。直到今天，我也不认为有准真的理解了超对称意味着什么，它对自然说了些什么——假如它正确的话。

多年过去了，我想我终于可以明白地说那时到底是什么驱使我脱离超引力的。我读爱因斯坦的原著学物理，对渗透到革命性的新物理学统一的那种思想有一种特别的体验。我希望的新统一应该从一个深层的原理出发，就像惯性原理或等效原理。从这一点看问题，我们能获得深刻而令人惊奇的洞察，发现原来不相干的两个事物在根本上是同一样东西。能量就是质量，运动与静止不可区分，加速度与引力是一样的。

超引力做的可不是这些。尽管它确实是一种新的统一构想，却只能靠令人心力交瘁的计算来表达和检验。我能做那些数学，但那不是我从爱因斯坦和其他大师的著作里学会的做科学的方法。

我那时结识的另一个朋友是斯特尔（Kellog　Stelle），他比我大几岁，也是德赛的学生。他们在一起研究超引力是否比广义相对论更容易与量子理论结合。由于背景独立方法还没有什么进展，他们也和别人一样用背景相关方法，虽然那方法在用于广义相对论时已经遭遇了惨重的失败。他们很快发现，它对超引力是很有成效的。他们检查了量子广义相对论首先出现无穷大的地方，看到了有限的结果。

这是好消息：超对称真的扭转了局面！可高兴没能长久。德赛和斯特尔只用几个月的时间就向自己证明了，超引力里的无穷大是随处可见的。实际的计算太难，几个月也算不好。但他们找到了一个办法，可以检验最终结果是有限还是无限。原来，所有的结果——比他们能直接检验的那个有限结果精确得多——都是无穷的。

然而事情还没完呢，因为还有其他形式的超引力等着检验。也许其中的某个能最终给出一个和谐的量子理论。每个形式都逐一经过了研究，都多少是有限的，因此还需要进一步延伸近似的序列，直到检验失败。虽然计算很难，但也没有理由认为有什么答案在经过了那个序列后就一定是有限的。一点渺茫的希望寄托在著名的N＝8理论，它也许与众不同。那最终是在巴黎构建起来的。但它也失败了——尽管还有人对它抱有希望。

超引力过去是、现在仍然是一个神奇的理论，但单凭它还不足以解决量子引力的问题。

于是，直到20世纪80年代初，构建量子引力理论的工作还没有丝毫进步。人们所有的努力（包括超引力）都失败了。规范理论胜利了，而量子引力的领域却停滞不前。我们还在为量子引力担心的少数几个人，感觉就像高中都没念完的学生，被请去参加妹妹在哈佛的毕业典礼，眼睁睁看着她同时获得医学、神经学和古印度舞蹈史的学位。

如果说超引力没能引出一个好的量子引力理论令我们沮丧，但也解放了我们。所有容易的事情都做过了。几十年来，我们通过推广费曼和他朋友的方法来构造理论。现在只有两件事情要做：放弃以固定背景几何为基础的方法，或者放弃认为在背景几何下运动的东西是点粒子的观点。两种方法后来都经过了探索，也都在通向量子引力的道路上——第一次——产生了巨大的成功。

# 第二篇　弦论简史

## 第七章　孕育革命

当我们遭遇一个不能仅仅以理解它的方式来解决的问题时，科学前进的步伐就会停下来。那是因为我们遗失了一样根本的东西，一种不同的技巧。不论我们多么勤奋，总不能找到答案，直到有人突然发现了那失去的链条。

人类第一次面对的这种事情大概是日食。经历过天空突然黑下来的一幕之后，天文学家们要做的第一件事情肯定是寻求一种方法来预言这种可怕的事件。人们在几千年前就开始了日食观测，同时还记录太阳、月亮和行星的运动。不久他们就明白J太阳和月亮的运动是周期性的，我们有证据证明人类在洞穴时代就知道那些事实了。不过日食要困难一些。

早期的天文学家对几件事情是很清楚的。太阳和月亮在空间沿不同的路线运行，日食发生在它们相遇的时刻。它们的路线在两处相交。只有当日月在那两处之一相遇时，才可能发生日食。因此，为了预言日食，必须跟踪太阳在一年的路径和月亮在一月的路径。只要跟踪两条路径，关注两个天体什么时候相遇，就可以预言了。其意义在于，必然存在一种在29天半的时间间隔内不断重复的模式，那就是月亮的周期。

但这个简单的思想是不对的：日食并不服从月亮周期所决定的模式。我们很容易理解前辈天文学家们的作为，他们想协调两大天体的运动，然而失败了。那对他们来说，也许和我们今天协调广义相对论

与量子论是一样的疑惑。

我们不知道是谁发现有元素丢失了，但不管是谁，我们都要感谢他。我们可以想象，一个天文学家，也许在巴比伦或古埃及，突然意识到原来需要考虑的周期运动不仅仅只有两个，而是三个。他也许是一个智者，经过几十年研究之后，就把数据都铭记在心了。他也许是一个年轻的叛逆，不囿于一定要用看得见的东西来解释看见的东西。不管情形怎样，他从数据中解开了神秘的第三个周期运动，它不是一年或一月发生一次，而是十八又三分之二年发生一次。结果发现，两个路径在天空的交点不是固定的：交点也在转动，需要十八年多的时间才转一圈。

第三个运动的发现——那丢失的元素——肯定算是抽象思维的一个最古老的成功例子。我们看到两个物体，太阳和月亮，很早就知道它们每个都有周期。而要“看到”还有一样东西也在运动，就需要想象了：它们的路径也在转动。这是深远的一步，因为它需要我们认识到在观测的运动背后还存在另一种运动，而那只能演绎推理才能发现。那时以来，科学通过发现那种丢失的环节取得进步的例子，只有寥寥的几个。

另一个这样的例子是人们发现基本粒子不是点粒子而是弦的振动。这对几个物理学大问题提供了合理的回答。如果它是对的，那么它与古人的圆周轨道本身也在运动的发现，是同等重要的认识。

弦理论的出现被称为科学革命，但它已经酝酿很长时间了。和有些政治革命一样——但不像过去的科学革命——有几个先驱者早就预言了弦理论革命的到来，他们在相对隔离的环境下奋斗了多年。他们从20世纪60年代开始研究强相互作用粒子——即夸克构成的粒子，如质子和中子，因而受强核力的作用——在散射时会发生什么。这不属于那五大问题，因为现在，至少在原则上，我们可以用标准模型来解释。但在标准模型之前，这是基本粒子物理学家的核心问题之一。

除了质子和中子，还有很多夸克组成的其他粒子，它们是不稳定的，是在加速器中打碎高能质子流而产生的。从20世纪30年代到60年代，我们积累了大量关于不同强相互作用粒子及其碰撞现象的数据。

1968年，年轻的意大利物理学家维尼齐亚诺（Gabriele Veneziano）从数据中发现了一种有趣的模式。他找到一个公式来描述这种模式，公式表达的是两个粒子相互散射时在不同角度出现的概率。维尼齐亚诺的公式惊人地符合一些实验数据。[35]

公式吸引了他在欧美的一些同事的兴趣，他们都对它感到疑惑。到1970年，有几个人已经可以用物理图像来解释公式了。根据那种图像，粒子不能看作点（它们以前总是被看作点的）；相反，它们更像“弦”，只存在于一维，可以像橡皮筋那样拉伸。它们获得能量时伸展，失去能量时收缩——也和橡皮筋一样。而且，它们也和橡皮筋一样振动。

维尼齐亚诺的公式就这样成了通向新奇世界的一道门，那个世界的强相互作用粒子都是橡皮筋，在运动中振动，彼此碰撞并交换能量。振动的不同状态对应着在质子破碎实验中产生的不同类型的粒子。

维尼齐亚诺公式的解释是芝加哥大学的南部阳一郎（Yoichiro Nambu）、玻尔研究所的尼尔森（Holger Nielsen）和斯坦福大学的苏斯金（Leonard Susskind）独立发现的。每个人都认为他做了一件迷人的事情，但发现他们的工作却没多大意思。苏斯金的文章被《物理学评论通讯》拒绝了，说他的见解还达不到发表的要求。后来，他在一次访谈中说，“嘭！我就像被垃圾筒打中了脑袋，感到非常非常憋屈。”[36]

但还是有几个人接受了它，并开始做研究。也许应该更准确地称

后来的思想为皮筋论。可那个名字有失尊严，所以诞生的是弦论。

作为强相互作用粒子的理论，弦理论后来曾一度被标准模型取代。但这并不意味着弦理论家错了；实际上，强相互作用粒子确实很像弦。我们在第四章讨论过，夸克之间的力现在由规范场描述，其基本定律由量子色动力学（QCD）确定，那是标准模型的组成部分。但在某些情形下，结果也可以描述为夸克之间连着橡皮筋。这是因为强核力与电磁力截然不同。电磁力随着距离增大而衰减，而夸克之间的力则在夸克分开时趋于常数，然后不论距离分开多远都保持那个常数。正因为这一点，我们在加速器实验中看不到自由夸克，而只能看到夸克组成的粒子。然而，当夸克靠近时，它们之间的力会减弱。这一点很重要。只有当夸克相距足够远时，才能满足弦（或皮筋）的图景。

最早的弦理论家们缺乏这个基本认识。他们想象了一个由橡皮筋周期性地连接的夸克世界——就是说，他们想让弦论成为一个基础理论，而不是任何更深层理论的近似。当他们想通过弦来理解弦时，麻烦就来了。问题来自他们为理论加的两个合理要求：首先，弦论应该与爱因斯坦的狭义相对论一致——就是说，它应该满足运动的相对性和光速的不变性。第二，它应该与量子理论一致。

经过几年的研究，人们发现弦论作为一个基础理论，只有在满足几个条件时才能与狭义相对论和量子论一致。第一，世界必须有25个空间维。第二，应该存在比光还快的粒子——快子。第三，应该存在不能静止的粒子。我们称这些粒子为无质量粒子，因为质量是静止粒子的能量的度量。

世界似乎没有25个空间维，为什么那理论没有被抛弃，这成了科学的一大疑问。我们能肯定的一点是，由于弦理论对额外空间维的依赖，在1984年前，很多人都没把它当真。很多人在观望，看到底谁是对的——是那些在1984年前拒绝多维的人还是那些在后来相信存在多

维的人？

快子也引出了问题。人们从未见过它们；更糟糕的是，它们的存在意味着理论是不稳定的，而且很可能存在矛盾。而且，在这种情形下，没有一个强相互作用粒子是零质量的，因此它不能作为强相互作用粒子的理论。

还有第四个问题。弦论包含着粒子，但不是所有自然存在的粒子。它没有费米子——也就没有夸克。这对想成为强相互作用的理论来说简直是一个巨大的难题！

四个问题中的三个是同步解决的。1970年，理论家拉蒙德（Pierre Ramond）改写了描述弦的方程，使它有了费米子。[37]他发现理论只有在具有新对称的情形下才能是和谐的。那种对称将混合新旧粒子——就是说，混合玻色子与费米子。拉蒙德就这样发现了超对称性；于是，不论弦论的命运如何，它是发现超对称的一条路线，也是孕育新思想的温床，已经硕果累累了。

新的超对称弦论还解决了两个其他问题。它没有快子，清除了人们接受它的一个主要障碍；它也没有25维，只有9个。虽然9维不是3维，但接近了很多。加上时间维，新的超对称弦（简称超弦）居于一个10维的世界。它比11少1，而奇怪的是，11是能写出超引力理论的最大维数。

大约同时，纳维（Andrei Neveu）和施瓦兹（John Schwarz）提出将费米子引入弦的第二种方法。和拉蒙德的理论一样，他们的理论也没有快子，也居于9个空间维的世界。纳维和施瓦兹还发现，他们可以让超弦发生相互作用，从而得到了与量子力学和狭义相对论一致的公式。

于是，只剩下一个疑难了。假如新的超对称理论包含零质量粒

子，那它如何成为强相互作用的理论呢？但事实上真的存在没有质量的玻色子。光子就是一个。光子永远不会静止，只能以光速运动。所以它有能量但没有质量。假想的与引力波相伴随的引力子也是这样的。1972年，纳维和另一个法国物理学家谢尔克（Joel Scherk）发现，超弦具有对应于规范玻色子（包括光子）的振动状态。这是朝正确方向迈出的第一步。38

两年后，谢尔克和施瓦兹迈出了更大的一步。他们发现，理论所预言的某些零质量粒子其实就是引力子。39（日本物理学家民秋米谷（Tamiaki Yoneya）也独立发现了同样的思想。）40

弦论包含规范玻色子和引力子的事实，令一切都改变了。谢尔克和施瓦兹马上就提出，弦论不是强相互作用的理论，而是一个基本理论——统一引力与其他力的理论。为了说明这是多么美妙而简单的思想，我们来看光子和引力子是如何从弦产生出来的。弦可以是闭的，也可以是开的。闭弦是一个圈，开弦是一根线，有端点。可能是光子的零质量粒子既来自开弦的振动，也来自闭弦的振动。引力子则只能来自闭弦（圈）的振动。

开弦的端点可以视为带荷的粒子。例如，一端可以是带负荷的电子，另一端可以是带相反电荷的正电子。端点之间的弦的无质量振动描述了在两个粒子之间传递电力的光子。于是，我们可以用相同的方式从开弦得到粒子和力。如果理论设计足够巧妙，它还可以生成标准模型里的所有力和粒子。

如果只有开弦，就没有引力子，这就把引力排除在外了。但人们发现还必须考虑闭弦，原因是大自然在粒子和反粒子之间产生碰撞，碰撞生成或湮灭光子。从弦的观点看，这可以描述为弦的两个端点靠近并结合在一起。端点消失了，留下一根闭弦。

实际上，为了让理论与相对论一致，粒子-反粒子湮灭与弦的靠近

是必须的，这意味着理论要求具有闭弦和开弦。但这说明它必须包括引力，而用开弦与闭弦的差别可以自然地解释引力与其他力的差别。这样，引力第一次在力的统一中扮演了中心角色。

这难道不美吗？它那么令人信服地把引力包括进来，任何有理性和知识的人——特别是那些在统一力的道路上追寻了多年而一无所获的人——大概仅凭这一点就会相信它，而不在乎它是否有具体的实验证据。

可它是怎么产生的呢？有什么定律要求弦的端点靠近并结合吗？这儿藏着理论最美妙的特征，运动与力达成了某种统一。

在多数理论中，粒子运动与基本力是两个不相干的事情。运动定律讲的是粒子在没有外力情形下的运动。从逻辑上讲，运动定律与力的定律没有关系。

在弦论中，情形就大不相同了。运动定律规定着力的定律。这是因为弦论中的所有力都有同样简单的起源——来自弦的分离与结合。一旦描述了弦如何自由运动，为了加入力的作用，我们只需要添加一根弦分裂为两根弦的几率。让这个过程时间反演，我们就能重新把弦结合起来（图7-1）。为了与狭义相对论和量子理论一致，弦的分裂与结合的法则终归是要预先严格规定的。力与运动就这样以一种奇特的方式统一了，这在点粒子理论中是不可想象的。

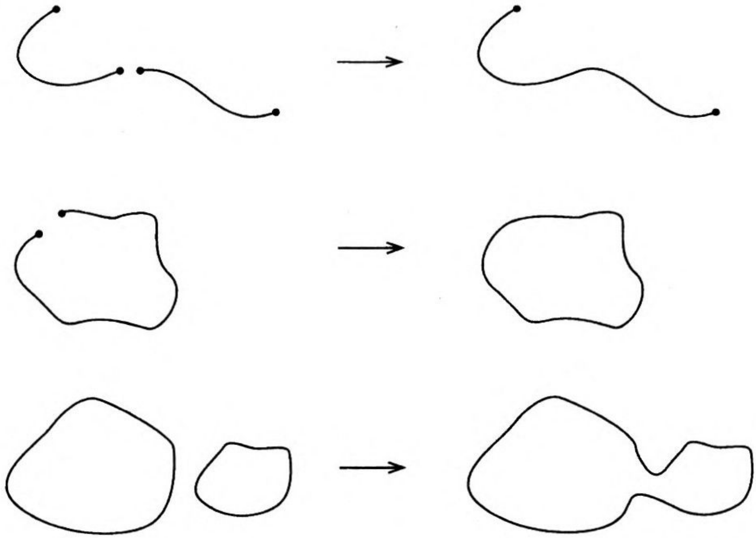


图7-1　上：两根开弦在端点结合；中：开弦的两个端点结合形成闭弦；下：两根闭弦结合成一根闭弦

力与运动的这种统一有一个简单的结果。在粒子理论中，可以自由加入所有类型的力，所以描述力的行为的常数可以无限增多。但在弦论中，只能有两个基本常数。一个叫弦张力，描述单位长的弦包含了多少能量。另一个叫弦耦合常数，描述一根弦分裂为两根弦——从而生成一个力——的几率；因为是几率，所以是一个数，没有单位。物理学的所有其他常数都必须与这两个数发生联系。例如，牛顿的引力常数原来是这两个数的乘积。

实际上，弦耦合常数不是自由常数，而是一个物理自由度。它的数值依赖于理论的解，所以不是决定定律的参数，而是标志解的参数。我们可以说一根弦分裂与结合的几率不是理论决定的，而是弦的环境决定的——就是说，是由它所在的多维世界决定的。（常数从理论的性质转移为环境的性质，这是弦论的一个重要方面，我们在下一章再讲。）除了这些而外，弦论满足的另一个定律既美妙也简单。想象吹肥皂泡，它会膨胀为一个完美的球形。下次你洗澡的时候可以留

心看看那些肥皂泡。它们的形状体现了一个简单定律，我们称它为肥皂泡定律。定律说，在一定的力和约束作用下，气泡的表面具有尽可能小的面积。

结果表明，这个原理同样适用于弦论。当一维弦在时间中运动，它就在时空中形成一个二维曲面（图7-2）。这个曲面有一定的面积，大致定义为弦长与其持续时间的乘积。

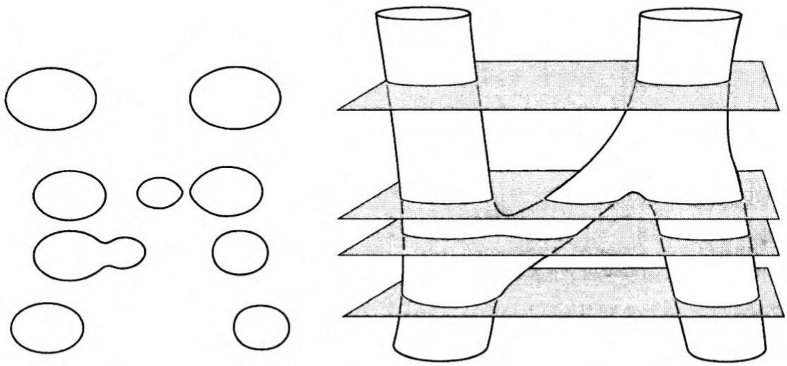


图7-2 弦的传播和相互作用取决于相同的定律，使其在时空中的表面积最小。在右边我们看到，两根闭弦在时空中运动。通过交换第三根闭弦发生相互作用。在左边，我们看到时空中的系列构型。是从右边的时空图中截取的碎片。我们先看到两根闭弦，然后看到第三根闭弦分裂出来，在运动中与另一根结合

弦运动使它在时空的面积最小，这就是全部的定律。它解释了弦的运动，而且，在弦能分裂或结合时，也解释了所有的力的存在。它用粒子的描述统一了我们知道的所有的力。它比它所统一的任何事物的定律都更简单。

弦论还达成了另一个统一。19世纪初，法拉第凭力线——磁极或正负电荷之间的线——想象电磁场的状态。对法拉第来说，这些力线是真实的；它们携带着磁极或电荷间的力。

在麦克斯韦理论中，场线成为场的次要特征，但也不一定非得如此。我们可以想象场线确实存在，粒子之间的力就是在它们之间延伸的场线。这在经典理论中不可能实现，但在量子论中却可以。

在超导体（即电阻很小或没有电阻的材料）中，磁场的场线成为离散的。每根线携带着某个极小量的磁流。我们可以将这些场线视为磁场的某种原子。20世纪70年代初，三个幻想家提出QCD的力线也是这样的，它们类似于电磁场的电力线。丹麦物理学家尼尔森就因为这个成为弦论的发明者之一——他将弦视为电流的量子化线。康奈尔的威尔逊（Kenneth Wilson）进一步发展了这幅图景，从此这种量子化的电场线被称作威尔逊线。第三个幻想家是俄罗斯物理学家波利亚柯夫（Alexander Polyakov），也许是思考规范理论与弦理论关系的最深刻的思想家。我读研究生时，听过他的一个发人深省的讲座，他宣布了他的一个宏愿，要把QCD重新表述为一个弦理论——而弦正好就是量子化的电流。

根据这些幻想家们的思想，规范理论中的基本对象是场线。场线满足简单的定律，定律规定了场线如何在电荷之间延伸。场本身是作为另一种描述而出现的。这种思维方式自然适合弦论，因为场线可以看做弦。

这意味着描述的对偶性：我们可以将场线视为基本的对象，而将定律视为对场线延伸和运动的描述；或者，我们可以将场视为基本对象，而将场线视为描述场的一种简便方式。在最子理论中，两种描述都是有效的。这引出一个原理，我们称它为弦和场的对偶性。不论哪种描述都行，都可以认为是基本的。

1976年，拉蒙德失去了在耶鲁的职位，而几年前他才解决了弦论的几个中心问题。看来，提出把费米子纳入弦论的方法、发现超对称、清除快子——哪怕是一鼓作气完成的——都不足以征服他的同事们，为他在常春藤盟校赢得一个教授的位置。

而施瓦兹在1972年就失去了普林斯顿的职位，尽管他对弦论有过他的基本的贡献。后来他到了加州理工学院，在经常变动的临时基金的资助下做了12年研究助理。只要愿意，他可以不讲课——可他也没有固定的职位。他发现了第一个能统一引力与其他力的好思想，但学院显然不相信他能进入正式员工的行列。

无疑，弦论的创立者们为他们的先驱性发现付出了很大的代价。为了认识这是怎样的一群人物，读者需要真正了解他们的工作的意义。当年一起读研究生的朋友已经是永久的正教授了，他们有很高的薪水，工作有保障，一家人衣食无忧。他们在著名的大学里有显赫的地位。而你呢，一无所有。你在内心深处知道他们走了捷径，而你做的是可能有着更大意义的事情，它们需要更多的创造力和更大的勇气。他们随波逐流，而你发现了崭新的理论。但你还是一个博士后，研究助理，或者毫无资历的教授。你没有长期的研究保障，前景渺茫。而你作为一个科学家，还是比别人更积极——发表更多的文章，带更多的学生——远远超过了那些做低风险研究却更有保障的人们。

那么，读者们，问问你自己，在这样的状况下，你想做什么？

施瓦兹坚持做弦论，继续寻找证据证明它可以是物理学的一个统一理论。尽管他还不能证明理论在数学上是和谐的，[41]但他确信他已经把握了一些东西。当第一代弦理论家们面临着难以逾越的障碍时，他们还在激励自己，假如基本粒子是弦的振动，那么所有的疑难都将得到解决。他们列出了一张令人兴奋的清单：

1.弦论为我们自动统一了所有的基本粒子，也统一了所有的力。它们都源自一种基本物体的振动。

2.弦论自动产生了规范场，它们决定着电磁力和核力。这些都自然源于开弦的振动。

3.弦论自动产生了引力子，它源自闭弦的振动；弦的任何量子理

论都必须包含闭弦。结果，我们自然得到了引力与其他力的统一。

4.超对称弦论统一了玻色子与费米子，两者都是弦的振动；因而所有的力与粒子也统一起来了。

而且，即使弦论不对，超对称性也仍然可以是对的。弦论比普通的量子理论更合适作为超对称性的依托。虽然超对称形式的标准模型丑陋而复杂，超对称的弦论却是美妙的珍玩。

最重要的是，弦论不费气力就实现了运动定律与力的定律的自然统一。

于是，弦论似乎达成了我们的一个梦想。整个标准模型，连同它的12种夸克和轻子，还有三种力外加引力，都能统一起来了，因为所有这些现象源自在时空延展的弦的振动，满足最简单的定律：弦运动时经过的表面积最小。标准模型的所有常数都可以归结为牛顿引力常数和一个代表弦分裂或结合的几率的简单数字的组合。即使那个几率的数不是基本的常数，却代表了环境的性质。

弦论承诺了那么多，难怪施瓦兹和他的几个合作者相信它一定是正确的。就统一问题而言，没有别的理论能在一个简单思想的基础上实现那么多东西。面对这样的前景，就只剩下两个问题了：它有效吗？需要什么代价？

1983年，我还是普林斯顿高等研究院的博士后，施瓦兹应邀来普林斯顿大学做了两个弦论的演讲。我以前没听过多少弦理论的东西，对他的演讲我只记得听众们紧张而不安的反应，一半是兴趣，一半是怀疑。威藤（Edward Witten）那时已经是基本粒子物理学的大人物，经常打断施瓦兹的演讲，问了他一连串长久的难题。我原以为这是怀疑的表现。后来我才逐渐明白，那是强烈兴趣的流露。施瓦兹很自信，但也有一点儿倔犟。我的印象是，他花了很多年的时间去感染别人对弦论发生兴趣。他的讲话使我相信他是一个勇敢的科学家，但

没能说服我做弦论的研究。那时，我认识的每个人都在做着自己的项目，而对那个新理论漠不关心。几乎没人意识到我们正生活在我们常说的物理学的末日。

## 第八章　第一次超弦革命

第一次超弦革命发生在1984年秋。称它为革命似乎有点儿自命不凡，但也名副其实。6个月前，只有少数大胆的物理学家在做弦论，除了个别同事，别人并不在意他们。正如施瓦兹说的，他和新合作伙伴、英国物理学家格林（Michael Green）“发表了好几篇论文，每次我都对结果感到激动……每次我们都觉得大家这回该对它发生兴趣了，因为他们可以看到这个学科是多么令人振奋。但还是无人喝彩。”[42]6个月后，几个最坚决的批评者也开始做弦论了。在新的氛围下，不丢弃自己正在做的事情而去追随他们，是需要勇气的。

事情的转折在于施瓦兹和格林的一个计算，它强有力地证明了弦论是一个有限而和谐的理论。更准确地说，他们终于成功地证明了，在超对称弦论（至少在10维时空）中，困惑许多统一理论的某些危险的方法（即所谓的“反常”）没有了。[43]我还记得对那篇论文的反应是既震惊又欢欣：震惊是因为有人怀疑弦论能在任何水平与量子论和谐一致；欢欣是因为格林和施瓦兹打消了怀疑，让人们期待统一物理学的最终理论已经握在我们手上了。

再也找不出更快的变化了。正如施瓦兹回忆的，就在我们要写完的时候，我们接到威藤的电话，说他听说……我们已经有了清除反常的结果。他想看看我们的工作。于是我们写了一个草稿，通过FedEx寄给他。那时没有email，它还没出现呢；但有了FedEx。所以我们寄给了他，他第二天就收到了。我们听说，普林斯顿大学和高等研究院的每一个人，所有的理论物理学家，都在做了，人数不少呢……于是，它在一夜之间就成了热门话题[笑]，至少在普林斯顿——很快就

影响了世界其他地方。这是很奇怪的事情，因为这么多年来我们发表了很多结果，却没人关心。可现在人们一下子变得兴奋起来。它从一个极端走到了另一个极端：从无人喝彩到万众欢呼……[44]

弦理论承担了其他理论没有的使命——它自诩为一个引力的量子理论，也是真正的力与物质的统一。它大胆而美妙的一击，似乎至少解决了那五个理论物理学问题中的三个。就这样，经过多年的失败之后，我们仿佛突然找到了黄金。（有趣的是，施瓦兹也马上从加州理工学院的高级研究助理提升为正教授。）

库恩（Thomas Kuhn）在他的名著《科学革命的结构》中为我们认为的科学历史上的革命性事件提出了一种新的认识路线。根据他的观点，科学革命发生之前往往出现大量的实验反常。结果，人们开始质疑现有的理论，还会有少数人另立新的理论。革命的高潮是实验结果有利于某个新理论而不利于旧理论。[45]库恩对科学的描述可能存在争议，我在本书的最后也会提出一些意见。但它描述了某些情况下发生的事件，还是有助于我们进行比较。

1984年的事情并不满足库恩的结构。现有的理论没有一个能解决弦论解决的问题。也没有实验反常；粒子物理学的标准模型和广义相对论足以解释那时的所有实验结果。即使如此，我们在一夜之间就有了一个可能的终极理论，能解释宇宙和我们在宇宙中的地位，又怎能不称其为革命呢？

1984年超弦革命四五年后，又有了许多进展，人们对弦论的兴趣也很快高涨起来。它成了最火爆的游戏。一头钻进来的人们都满怀着雄心和骄傲。有很多新技术等着他们去学，需要投入几个月甚至一年的时间才可能做弦论，这对一个理论物理学家来说是很漫长的时间。认真做的人看不起不做或不能做的（总会有这种想法的）。很快就形成了差不多和宗教派别一样的团体，你要么是弦理论家，要么不是。我们几个人还想坚持走普通路线：有了好想法，我会做下去，而同时

也会走别的路线。很难一条路走到头，因为站在潮头的那些人不大愿意搭理我们这些没有加入新潮的人。

为了适应新领域，很快就有了弦论的学术会议。那些会议仿佛都在欢呼胜利，好像已经真的发现了什么理论，而别的一切都无所谓，都不值得考虑。弦论的讲习班开始在主要的大学和研究机构流行。在哈佛，弦论讲习班被称为“后现代物理班”。

这个名称没有讽刺的意味。弦论的讲习班和学术会很少讨论的一个课题是，如何用实验来检验理论？虽然有几个人为此忧虑，其他人却认为那是不必要的。他们感觉只可能有一个能统一所有物理学的和谐的理论，而弦论看来做到了这一点，所以它一定是对的。检验我们的理论不再需要实验了，这是伽利略的精神。数学已经足以探索自然定律。我们走进了后现代物理学时代。

很快，物理学家意识到弦论不是唯一的。我们很快发现，在10维时空里有5个和谐的超弦理论，而不是只有一个。这引出一个难题，大约10年后才得到解决。不过，这还不算坏消息。我们还记得卡鲁扎-克莱因理论有一个致命问题：它描述的宇宙太对称了，不符合实际，因为从镜像看到的自然是不一样的。5个超弦理论能避免这个难题，它们描述的世界和我们真实的世界一样不那么对称。进一步的发展证明弦论是有限的（就是说，它关于任何实验结果的预言都是有限的数值）。在玻色弦而没有费米弦的情形，很容易证明没有类似于引力子理论中的那些无穷表达，但当概率计算到达某个更高的精度时，无穷大还是会出现，它与快子的不稳定性有关。因为超弦没有快子，于是理论可能没有无穷大。

在低阶近似下，这是很容易证明的。除此之外，直觉也告诉我们理论在每阶近似下都应该是有限的。我想起一个杰出的弦理论家说过，弦论的有限性实在太明显了，即使有那样的证明，他也不想去研究。但还是有人在努力证明弦论在最低阶的有限性。最后，1992年，

伯克利的一个德高望重的数学家曼德尔斯塔姆（Stanley Mandelstam）发表了一篇论文，人们相信它证明了超弦理论在一定近似策略下的每一阶都是有限的。[46]

难怪大家那么乐观。弦论的前景大大超越了从前的任何一个统一理论。同时我们可以看到，为了实现那些蓝图，还有漫长的路要走。例如，考虑标准模型常数的解释问题。正如上一章讲的，弦论只有一个常数可以人工调节。如果弦论是正确的，标准模型的20个常数必须用那一个常数来解释。假如这些常数在弦论中都能作为一个常数的函数进行计算，那将是无法用语言来形容的一个奇迹——是科学史上最伟大的胜利。但我们还没走到那一步。

此外还有一个问题——我们以前讨论过的，任何统一理论都存在那个问题。如何解释那些统一的粒子之间的显著差别呢？弦论统一了所有的粒子和力，意味着它必然也能解释它们为什么不同。

于是，理论要落实到细节上来了。它真的有效吗？还有令那奇迹失色的模糊的东西吗？假如它有效，那么简单的理论究竟如何解释那么多东西呢？如果弦论是对的，我们对自然该相信什么呢？在这个过程中我们有失去的东西吗？失去的是什么呢？

随着我对弦论的更多了解，我开始认为它所提出的挑战很像我们在买汽车时面临的难题。你带着一堆条件去车行，经销商很乐意把满足那些条件的车卖给你。他拿出几个模型，你突然意识到每辆车都有某些你不需要的功能。你想要铁锁刹车和好音质的CD播放机。具有那些功能的汽车还配备着阳蓬、特制的铬合金保险缸、钛钢的轮毂盖、8个杯座、特制的赛车条纹。

这就是大家熟悉的打包式的销售。事实证明，你不可能买到只有你需要的功能的汽车。你会得到一揽子的东西，包括你不想要的和不需要的。这些额外的东西提高了价格，但你别无选择。假如你想铁锁

刹车和CD机，你就必须把整套东西都拿走。

弦论似乎也只能做打包的买卖。你大概只想要一个统一所有粒子和力的理论，但你得到的还有额外的东西，至少有两样是不能讨价还价的。

第一个是超对称性。有的弦理论不具备超对称性，但已经知道它们都是不稳定的，因为存在讨厌的快子。超对称性似乎消除了快子，但也带来了麻烦。超对称弦论只有在具有9个空间维的宇宙中才可能是和谐的。在3维空间里不可能有那样的理论。假如你想要其他性质，你就必须接受那额外的6维。这一点引出了很多麻烦。如果不能立刻排除这个理论，就必须想办法把多余的维隐藏起来。这似乎别无选择，只能将它们卷起来，小到不能察觉的程度。于是我们不得不让旧的统一理论的主要思想重新复活。

这提供了好机会，也提出了大问题。我们看到，过去人们用高维来统一物理学的努力都失败了，因为问题的解太多；高维的引进产生了唯一性的大难题。它还引出了不稳定性问题，因为存在不同的过程，在有的过程中，额外的维会张开而变得很大，而在另一些过程中它们会坍缩而形成奇点。弦论要想成功，就必须解决这些问题。

弦理论家很快意识到唯一性问题是弦论的基本特征。现在有6个额外维需要卷曲，而且有很多卷曲方式。每种方式都涉及复杂的6维空间，都产生不同形式的弦理论。因为弦论是背景相关的理论，从技术角度说，我们对它的理解是它描述了在固定背景几何下运动的弦。通过选择不同的背景几何，我们得到技术上不同的理论。它们源自相同的思想，相同的定律适用于每一个情形。但严格说来，每一个都是不同的理论。

这不仅是鸡毛蒜皮的事情。不同理论给出的物理预言也是不同的。多数的6维空间由一系列常数来描述，它们可以自由确定。这些

常数标志着不同的几何特征，如额外维的体积等。一个典型的弦理论可以有几百个常数，它们是描述弦传播和相互作用的一部分。

考虑一个二维曲面的物体（如球面）。因为它是完全球形的，只需要一个参数（圆周长）来描述。现在考虑更复杂的曲面（如面包圈，图8-1）。这个曲面由两个参数描述。面包圈有两个圆周，沿不同的方向环绕它。两个圆周可以有不同的周长。

我们可以考虑更复杂的带有多个孔的曲面。它们需要更多的数来描述。但没有谁（至少我不知道）能直接把6维空间的形象画出来。

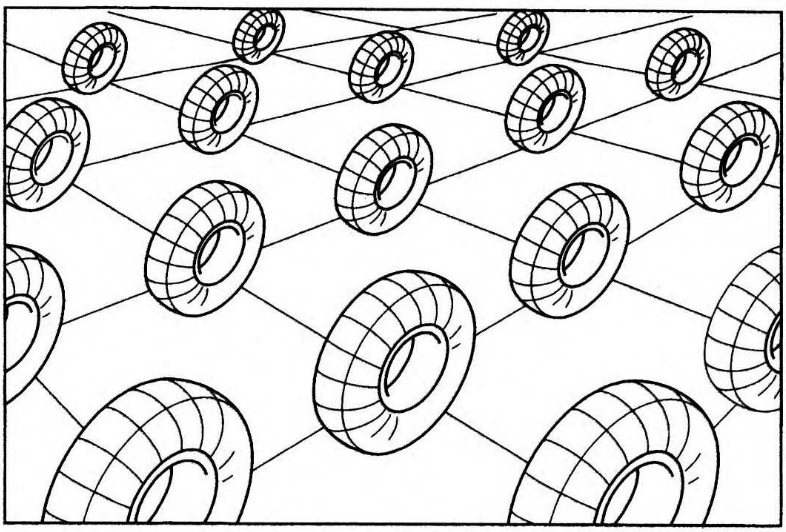


图8-1　隐藏的维可以有不同的拓扑。在这个例子中，有两个隐藏的空间维，它们具有和面包圈（环）一样的拓扑

然而，我们还是有描述它们的办法，就是类比面包圈或其他二维曲面可能出现多少个孔洞。我们不用弦来缠绕孔洞，而是用高维空间来缠绕它。在每种情形，缠绕的空间具有一定的体积，那将是描述那种几何的一个常数。当我们明白了弦如何在额外维中运动，所有常数都会出现。所以，我们不再只有一个常数，而是有很多常数。

弦论就这样克服了物理学的统一所面临的困境。即使一切都来自一个简单原理，我们也必须解释为什么会出现那么多不同的粒子和力。在最简单的可能情形，空间有9维，弦论很简单；所有相同类型的粒子都是一样的。但当弦可以在6个额外维空间的复杂几何中运动时，就将产生许多不同类型的粒子，伴随着在每个额外维的不同运动和振动方式。

于是我们为粒子之间的显著差别找到了自然的解释，这是统一理论必须做的事情。但这也付出了代价，即理论远不是唯一的。结果是常数发生了交换：标志粒子质量和力的强度的常数与刻画额外维几何的常数相互交换了。所以，找到能解释标准模型的常数也就不足为奇了。

即使如此，倘若这个图景能为标准模型的常数给出唯一的预言，它仍然是很吸引人的。如果我们通过将标准模型的常数转化为额外维几何的常数而发现了标准模型常数的某些新东西，如果这些发现与自然一致，那么这些证据将强有力地证明弦论是正确的。

但事实不是这样的。在标准模型里自由变化的常数转化为弦论中可以自由变化的几何。没有约束，也没有简化。因为额外维的几何有很多选择，自由常数的数目不是减少了，而是增多了。

而且，标准模型也没有完全重现。我们确实能导出它的一般特征，如费米子和规范场的存在，但不能从方程得到自然出现的复合现象。

事情从此越发恶劣了。所有弦理论都预言了额外的粒子——没有在自然看到过的粒子。伴随它们的还有额外的力。有些力来自额外维几何的变化。考虑在空间的每一点加一个球面，如图8-2。球的半径可以随我们在空间的运动而变化。

因此，每个球的半径可以作为它所在的点的一个性质。就是说，它像一个场。正如电磁场一样，这种场也在空间和时间中传播，生成额外的力。这一点很清楚，但这些额外的力很可能不会与我们的观测一致。

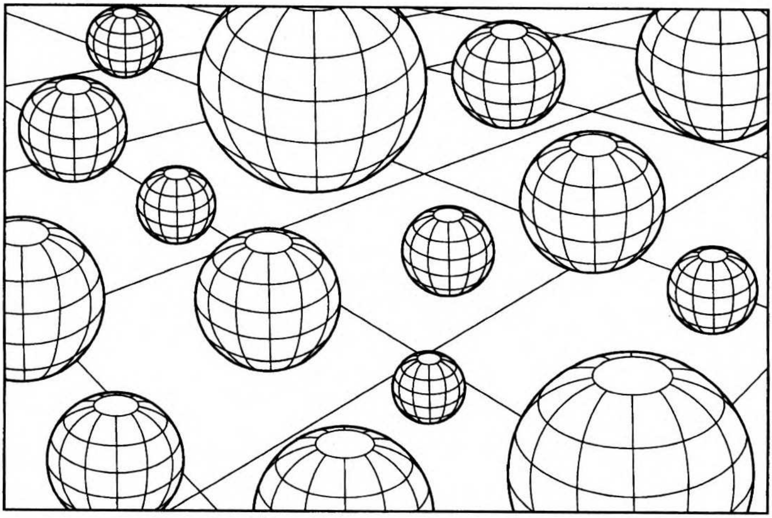


图8-2　隐藏维的几何可以在时空中变化。在这个例子中，球半径是变化的

我们一直在谈一般性，但世界只有一个。如果弦论是成功的，它不但应该是所有可能世界的模型，还必须解释我们的世界。于是，关键的问题是：有什么办法卷曲额外的6维，从而完全再现粒子物理学的标准模型？

有一种方法是找一个超对称的世界。因为弦论具有超对称性，于是，对称性在我们三维世界的具体表现将依赖于额外维的几何。我们可以想办法让超对称在我们的世界破缺。或者也可能是另一种情形，即现实的理论容不下许多的超对称性。

这就产生一个有趣的问题：是否可以选择某个额外6维的几何，使其正好满足超对称性的数量？是否可以构造某个几何，使我们的三

维世界具有超对称形式的标准模型所描述的某个粒子物理学？

这个问题在1985年的一篇非常重要的论文得到了解决。文章的作者是四个弦理论家：坎德拉斯（Philip Candelas）、霍洛维茨（Gary Horowitz）、斯特罗明戈（Andrew Strominger）和威藤。[47]他们很幸运，因为数学家卡拉比（Eugenio Calabi）和丘成桐（S-hing-Tung Yau）已经解决了一个给他们带来答案的数学问题。两个数学家发现并研究了一种特别优美的6维几何形式，即现在所谓的卡丘空间。四个弦理论家证明，弦论实现某个超对称标准模型的条件也就是确定卡丘空间的条件。接着他们提出，描述大自然的弦论选择了卡丘空间作为额外6维的几何。这清除了许多其他的可能性，给理论赋予了更多的结构。例如，他们具体说明了如何将标准模型中的常数（如决定不同粒子质量的常数）转换为描述卡丘空间几何的常数。

这是一大进步，但也同样存在一大问题。假如只有一种卡丘空间，具有固定常数，那么我们就得到了渴望已久的唯一的统一理论。遗憾的是，竟然有很多卡丘空间，没人知道有多少。丘本人说至少有10万个。每个空间都产生不同形式的粒子物理；每个空间都伴随着一连串决定空间形状和大小的自由常数。所以没有唯一性，没有新预言，也没解释任何东西。

另外，涉及卡丘空间的理论还有很多额外的力。结果表明，只要弦理论是超对称的，许多力都将具有无限的作用范围。这是很不幸的，因为任何作用范围无限的力（而不仅是引力和电磁力）的存在，都有严格的实验限制。

还有一个问题。决定额外维几何的常数可以连续变化。这可能引出像旧的卡鲁扎-克莱因理论那样的不稳定性。除非有什么神秘的机制能固定额外维的几何，这些不稳定性将导致灾难，例如额外维的坍缩会产生奇点。

另外，即使我们的世界由一种卡丘空间的几何描述，也不能解释它是如何成为那样的。除了卡丘空间外，弦理论也有多种形式。在有的形式中，卷曲维的数目从零一直变到9。我们称那些没有卷曲维的几何叫平直的，它们确定了我们这些大生物经历的世界。（在考察它们对粒子物理学的意义时，我们可以忽略引力和宇宙学，在那种情形，非卷曲维具有狭义相对论所描述的几何。）

10万个卡丘流形不过是冰山的一角。1986年，斯特罗明戈发现了一个方法，能构造大量其他形式的超对称弦理论。他在描述他的构造方法的论文的结论部分写了一段话，我们应该好好记在心里：

这类超对称的超弦紧化空间大大地增加了……这些解似乎……不能在可见的将来得到分类。这些解的约束相对较弱，也许可能找到大量在现象上可以接受的解……虽然这令人充满了信心，但从某种意义说，生命也变得太简单了。所有的预言能力似乎都消失了。（我强调的）

所有这些都表明我们当前最最需要的是找一个动力学原理来决定[哪个理论描述了自然]，现在它比已往任何时候都更加急迫。48

于是，弦论在采用过去的高维理论的策略时，也给自己带来了问题。有很多解，有些解大概能描述真实的世界，但大多数解却不能。还有很多不稳定性，表现为大量额外的粒子和力。

这注定会引发争议，而且争议确实来了。很少有人能否定它有很多好的令人满意的特征。粒子是弦的振动的思想，也的确像是丢失了的那个环节，能解决很多未解的问题。但代价也很高昂。我们被迫接受的那些打包的特征也损害了理论原先的美——至少对我们几个人来说。在其他人看来，额外维的几何是弦论最美妙的东西。难怪理论家们泾渭分明地成了两派。

相信弦论的人往往也相信它的整个一揽子的特征。我认识许多理

论家都确信超对称和额外维是存在的，只等着我们去发现。我也认识很多在这一点上脱离弦论的人，因为它意味着要接受太多没有实验基础的东西。

费曼是个典型的看不起弦论的人，他曾解释过为什么不愿赶他们的潮流：

我不喜欢他们不做任何计算。我不喜欢他们不检验自己的思想，我不喜欢他们为任何不符合实验的东西虚构解释——一句自我安慰的话就是，“啊，它仍然可能是对的。”例如，理论需要10维。好啊，也许有办法把多余的6维卷起来。是的，那在数学上是可能的，可为什么不是7呢？当他们写方程时，方程就应该决定多少东西卷曲了，而不是为了迎合实验才要求那样。换句话说，在超弦理论中没有任何理由能说明为什么卷曲的不是10维中的8维而只剩下2维，那当然完全不符合我们的经验。所以，即使它可能与经验不符也无关紧要，其实它什么结果也没有；它只能浪费时间。它看起来就不对。[49]

很多老一辈的粒子物理学家都带有这种情绪，他们知道粒子物理学的成功总是需要与实验物理学进行不断的交流。另一个反对者是因为标准模型获诺贝尔奖的格拉肖：

但超弦物理学家还没证明他们的理论确实有效。他们不能证明标准模型是弦理论的逻辑结果。他们甚至不能肯定他们的体系包括了诸如质子和电子的描述。他们连一丁点儿的实验预言也没拿出来。最糟糕的是，超弦理论不是根据一组迷人的关于自然的假定而得到的逻辑结论。你大概要问，为什么弦理论家坚持空间是9维的？只因为弦理论在其他任何空间都没有意义……[50]

然而，争议之外还需要更好地认识理论。一个理论以那么多不同的面目出现，就不像是一个单独的理论了。如果说这些理论还算什么东西的话，它们似乎是另外某个未知理论的不同的解。

我们习惯了一个理论有多个不同的解的观念。牛顿定律描述粒子如何在外力作用下运动。如果我们把力固定——例如，我们想描述在地球重力场中抛出的球，牛顿方程具有无限多个解，对应于球的无限多个可能的路径：它可高可低，可快可慢。每种抛球的方法都产生一个不同的路径，每个路径都是牛顿方程的一个解。

广义相对论也有无限多个不同的解，每个解都是一个时空——也就是宇宙的一个可能的历史。因为时空几何是动力学实体，它可以存在于无限多种不同的构形，演化成无限多个不同的宇宙。

弦理论的每个背景都定义为爱因斯坦方程或其某种推广形式的解。于是，在人们看来，名目越来越多的弦理论意味着我们并不是真的在研究一个基础理论。我们只不过在研究某个更深层理论——一个我们还不知道的理论——的解。我们也许可以称那个理论为元理论（meta-theory），因为它的每个解都是一个理论。那个元理论才是真正的基本定律。它的每个解都将生成一个弦理论。

这样看来，更引人入胜的是，我们不去考虑无限多个弦理论，而是考虑从某个基础理论产生的无限多个解。

回想一下，每个弦理论都是背景相关的理论，它们描述在特殊的背景时空下运动的弦。因为不同的近似的弦理论处于不同的时空背景，一个理论若要统一它们，就一定不能处于任何时空背景。为了统一它们，我们需要一个单一的背景独立的理论。该怎么做，也就很清楚了：构造一个本身是背景独立的元理论，然后从那个元理论导出所有背景相关的弦理论。

于是我们有两个理由寻求背景独立的引力的量子理论。我们已经知道，我们必须融合爱因斯坦广义相对论所给出的几何的动力学特征。现在我们要用它来统一所有不同的弦理论。这需要新的思想，但至少现在我们还没有。

元理论要做的一件事情就是帮助我们选择哪种形式的弦理论能在物理上实现。因为大家都相信弦论是唯一的统一理论，许多理论家料想大多数形式的理论是不稳定的，而那个真正稳定的理论将唯一解释标准模型的常数。

80年代后期的某个时候，我突然想起还有另一种可能性。也许所有弦理论都是同样有效的。这将意味着我们对物理学的期待要彻底改变，即基本粒子的所有性质都将成为偶然的——不是基本定律决定的，而是基本理论的无限多个解中的某一个决定的。已经有证据表明这种偶然性可以伴随自发对称破缺而在理论中发生，但众多形式的弦理论让我们看到了新的可能——即从本质上说，偶然性也适用于基本粒子和力的所有性质。

这意味着基本粒子的性质是环境决定的，可以随时间而变化。如果是那样，它将意味着物理学更像生物学，即基本粒子的性质会依赖于我们宇宙的历史。弦论将不是一个理论，而是一幅理论构成的图景——就像进化生物学家研究的适应性景观。甚至可能存在某种像自然选择那样的过程，选择出适合我们宇宙的理论形式。（这些思想导致了1992年的一篇文章，题为“宇宙进化吗？”[51]，还有1997年的一本书，题为《宇宙的一生》。我们以后还要回来谈这些思想。）

每当我与弦理论家们讨论这种进化原理，他们总会说，“别担心，会有唯一形式的弦理论的，它将由某个我们至今尚未认识的原理来决定。当我们发现它时，这个原理将正确解释标准模型的所有参数，并对未来实验提出唯一的预言。”

不管怎么说，弦论的步伐慢下来了，到20世纪90年代初，弦理论家们也泄气了。弦理论没有完整的形式，我们有的只是一张罗列着几十万个不同理论的清单，而每个理论都有许多自由常数。这么多的理论中，哪些对应于现实，我们没有一点明确的概念。尽管技术进步了很多，但没有一个证据能告诉我们弦理论是对还是错。最糟糕的是，

它没有提出一个具体的能用目前实验来证明或证伪的预言。

弦论令人泄气还有别的原因。80年代末是整个领域的黄金时期。1984年革命刚过，弦论的创立者（如施瓦兹）们就收到了很多来自顶尖大学的诱人邀请。几年里，年轻的弦理论家成长起来了。但到90年代初，一切都过去了，有才能的人还是找不到工作。

有些人（不论老的还是年轻的）就在这个关头离开了弦论。幸运的是，弦论的工作是很好的智力训练，从前的弦理论家如今活跃在其他领域，如固体物理学、生物学、神经科学、计算机和金融业。

还有些人仍然坚守弦论的阵地。尽管有那么多泄气的理由，许多弦理论家还是念念不忘弦论构建了未来的物理学。如果说有什么问题，那是当然，统一基本粒子的其他方法也没有哪个成功的呀。还有少数人在做量子引力，不过多数弦理论家都假装不知道。在他们许多人看来，弦论才是唯一的选择。纵然长路漫漫，比他们想象的艰辛得多，但没有其他任何理论有望能在一个有限而和谐的框架下统一所有的粒子和力并解决量子引力。

不幸的结果是，信奉者和怀疑者分裂更远了。每一派都在加强自己的阵地，似乎都为坚守各自立场找到了好理由。如果不出现戏剧性的进步，极大改变我们对弦论的态度，这样的局面还会僵持很长时间。

## 第九章　第二次革命

弦论最初提出来，是为了统一自然的所有粒子和力。但经过1984年革命以来的10年研究，发生了意想不到的事情。这个本以为统一的理论分裂成了许多不同的理论：10维空间里的5个和谐的超弦理论，外加不同卷曲维下的几百万个不同形式的理论。十几年过去了，我们现在明白了弦论本身正需要统一。

第二次超弦革命发生在1995年，它正是那样的一场统一运动。革命的起因通常认为是那年3月威藤在洛杉矶弦论会议上的一个讲话，他提出了一个统一它们的设想。其实他并没有拿出一个新统一的超弦理论，而只是说存在那样的理论，它会有哪些特征。威藤的建议是基于最近的系列发现，它们揭示了弦论的一些新面目，极大增进了我们的理解。那些发现还揭示了规范理论和广义相对论之间更多的共性和联系，进而用它们统一了弦理论。这些进步（其中有的是现代理论物理学史上前所未有的）最终赢得了很多怀疑者，也包括我。起初，5个和谐的理论似乎描述了不同的世界，但到90年代中期，我们开始明白它们并不像表面那么不同。

如果出现了两种不同的方式来看同一个现象，我们就说它们有对偶性。分别让一对夫妻给你讲他们的故事，他们的说法会不同，但每个重要事件都能相互得到印证。和他们谈话多了，你就能指出两人说的故事有什么不同和联系。例如，丈夫觉得妻子过于自信，这正好印证了妻子抱怨丈夫太懦弱。我们可以说，两个人的话是互为对偶的。

弦理论家在寻找5个理论的相互关系时，开始运用不同类型的对偶性。有些对偶性是精确的：两个理论不是真的不同，只是描述同一现象的两种不同方式。其他对偶性是近似的，在这些情形，两个理论确实不同，但一个理论的现象类似于另一个理论的现象，这样就可以通过研究一个理论的某些特征来近似地了解另一个理论。

5个超弦理论中最简单的对偶性叫T对偶。“T”代表“拓扑的”，因为这种对偶性与空间的拓扑有关。[52]当某个紧化的空间维是圆时，就出现这种对偶。这时，弦可以缠绕在圆周上。实际上，它可以缠绕很多圈（图9-1）。弦缠绕圆周的圈数叫缠绕数。

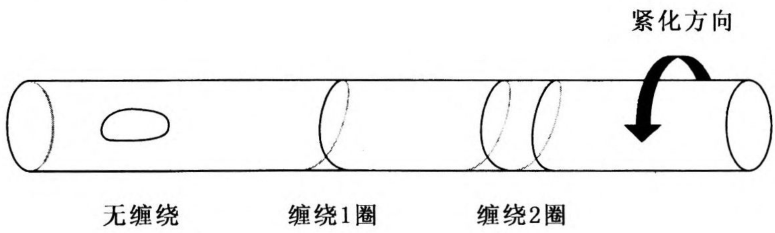



图9-1　弦可以缠绕一个隐藏维。在这里的情形，空间是1维的，隐藏维是个小圆。图中的弦分别缠绕零圈、一圈和两圈

另一个数度量弦如何振动。这种弦和钢琴或吉他的弦一样，也有泛音，可以用自然数标记不同振动的音阶。T对偶就是两个缠绕着圆的弦理论之间的关系。两个圆的半径不同，但相互关联；一个等于另一个的倒数（以弦长为单位）。在这种情形，一个弦理论的缠绕数完全表现为另一个弦理论的振动音阶。这种对偶性出现在5个弦理论的某些对之间。它们看似从不同的理论出发，但把它们的弦缠绕在圆圈上时，就成为同一个理论了。

还有第二类对偶，人们也猜想它是精确的，尽管还没有证明。我们在第七章讲过，每个弦理论都有一个数决定弦分裂或结合的几率。这是弦的耦合常数，约定以字母g标记。当g很小时，弦分裂或结合的几率就小，我们就说相互作用弱。当g很大时，弦随时都在分裂或结合，我们就说相互作用强。

于是，两个理论又可能以下面的方式发生联系：每个理论都有耦合常数g。但是，当一个理论的g等于另一个理论的1/g时，两个理论的表现就会是一样的。这叫S对偶（S代表strong-weak，强弱）。如果g小，意味着弦相互作用弱，而1/g大，所以另一个理论中的弦相互作用强。

耦合常数不同的两个弦理论怎么可能有相同的行为呢？难道我们连弦分裂或结合的几率是大还是小都说不准了吗？只要知道了弦是什

么，我们是能说清楚的。但事实是，在S对偶的情形下，我们相信两个理论拥有的弦比我们想象的更多。

弦的增生是一种常见但少有人认识的所谓“突现”现象的一个例子，“突现”一词所描述的是从巨大的复杂系统中生出新的性质。我们也许知道基本粒子满足的定律，但许多粒子束缚在一起时，各种新现象就会涌现出来。质子束、中子束和电子束可以结合生成新的金属；同样数目的其他东西可以结合生成生命的细胞。不论金属还是细胞，都不过是质子、中子和电子的集合体。那么，我们该如何来描述是什么让金属成为金属，细胞成为细胞的呢？区别二者的性质就叫突现性质。

看一个例子：金属最简单的行为大概就是振动；如果你敲击金属棒的一端，就会有声波从它穿过。金属振动的频率就是一种突现性质，声波在金属内传播的速度当然也是。想想量子力学中的波粒对偶，意思是每个波都伴随着一个粒子。反过来也是对的：每个粒子都伴随着一个波，也包括伴随着在金属中传播的声波的粒子，它叫声子。

声子不是基本粒子，当然也不是构成金属的粒子，因为它只能凭借构成金属的大量粒子的集合运动才能存在。但声子仍然还是粒子。它具有粒子的一切性质。它有质量，有动量，也携带能量。它的行为和量子力学规定的任何粒子应有的行为是一样的。我们说声子是突现粒子。

我们相信，弦也会发生这样的事情。当相互作用强时，有许许多多弦在分裂、结合，因而很难分辨哪根弦发生了什么。于是我们寻求大量弦的集合的某些简单的突现性质——通过那些性质来认识发生了什么。结果真的出现了有趣的事情。正如一束粒子的振动可以表现得像一个简单的粒子（声子），从大量弦的集合运动中也生出一根新弦，我们称它为突现弦。

突现弦的行为与普通的弦（我们不妨称其为基本弦）恰好相反。相互作用的基本弦越多，突现弦就越少。说得更准确一点儿：假如两根基本弦相互作用的几率正比于弦耦合常数g，那么在某些情形下，突现弦发生相互作用的几率就正比于1/g。

怎么区分基本弦与突现弦呢？事实证明区分不了——至少在某些情形是这样的。实际上，我们可以转换图像，把突现弦看作基本弦。那是强弱对偶性的一个奇异技巧。那就像我们在考虑金属时，把声子（声波的量子）看成基本的，而把构成金属的所有质子、中子和电子看成由声子构成的突现粒子。

和T对偶一样，这种强弱对偶也关联着5个弦理论中的某些对。唯一的问题是，这种关系仅适用于理论的某些状态抑或有着更深层的意义？这之所以成为问题，是因为我们必须研究某些理论对的状态——特定的对称性约束下的状态，才可能揭示那种关系。否则，我们就不能充分控制计算而得出好的结果。

接着，理论家们面临着两条可能的路线。乐观的一派——那时多数弦理论家都很乐观——走得很远，他们在证明结果的基础上，进而猜想他们在理论对中检验的特殊对称状态之间的关系，可以扩展到所有5个理论。就是说，他们假定即使没有特殊对称性，也总会存在突现弦，而那些突现弦也总是表现为其他理论中的基本弦。这意味着S对偶不仅联系着理论的某些方面，而且证明了它们的完全等价。

另一方面，少数悲观者担心5个弦理论也许真的彼此不同。在他们看来，哪怕只有在很少的情形下，一个理论的突现弦能像其他理论的基本弦，也是相当了不起的了。但他们意识到，这种事情即使在所有理论都不同的时候也可能是真的。

很多人曾观望（现在仍然在观望）乐观派与悲观派的对错。如果乐观派对了，那么原来的所有5个超弦理论都不过是同一个理论的不

同描述形式。如果悲观派对了，那么它们真是不同的理论，因而没有唯一性，没有基本理论。只要我们不知道强弱对偶是近似的还是精确的，我们就不能知道弦理论是不是唯一的。

支持乐观派观点的一个证据是，相似的对偶性也存在于比弦理论更简单也更容易理解的理论中。有一种形式的杨-米尔斯理论，所谓“N＝4超杨-米尔斯理论”，就是那样一个例子，它有着尽可能多的超对称性。为简单起见，我们将称它为最大超理论。有很好的证据表明这个理论具有某种形式的S对偶。它的行为大致是这样的：理论有大量带电荷的粒子，也有某些带磁荷的突现粒子。在通常情况下没有磁荷而只有磁极。每个磁体有两个磁极，分别叫南极和北极。但在特殊情况下，可以有彼此独立运动的磁极——它们就是著名的磁单极。最大超理论发生的情况是，存在某种电荷与磁单极交换的对称性。当两者交换时，如果将电荷值改变为原来数值的倒数，则理论描述的物理不会有任何改变。最大超理论是一个非同寻常的理论，我们很快就会看到，它将在第二次超弦革命中发挥巨大作用。不过，既然我们对不同的对偶性已经有了一点认识，我可以来解释威藤在洛杉矶的著名讲话中讨论的那个猜想了。

我说过，威藤讲话的关键思想是5个和谐的超弦理论其实是同一个理论。但那个单独的理论本来是什么呢？威藤没说，不过他确实描述了一个大胆的猜想，认为那个统一5个超弦理论的理论需要再多一维，这样空间就有10维，而时空是11维。[53]

这个特别的猜想首先是两个英国物理学家赫尔（Christopher Hull）和汤森（Paul Townsend）在一年前提出的。[54]人们已经发现，对偶性不仅存在于那5个理论中，也存在于任何弦理论以及11维的理论中，威藤在此基础上为猜想找到了大量的证据。

为什么弦理论的统一需要多1维呢？额外维的性质——卡鲁扎-克莱因理论中的圆半径——可以解释为在其他维上变化的场。威藤根据

这个类比提出，弦理论的某个场其实就是在第11维延展的那个圆周的半径。

引进1个额外的空间维有什么用呢？毕竟，在11维时空里没有一个和谐的超对称弦理论。但在11维时空里还真有那么一个超对称的引力理论。你大概还记得，第七章说过那是所有超引力理论中维数最高的一个，是超引力的真正的珠穆朗玛峰。所以，威藤猜想那个11维的世界（额外的场指示了它的存在），在没有量子理论的情形下，可以用11维超引力来描述。

而且，尽管11维里没有弦理论，但关于在11维时空中运动的2维曲面，却自有它的理论。那个理论至少在经典水平上是很美妙的。它是1980年代初出现的，有个富有想象的名字，叫11维超膜理论。

在威藤之前，多数弦理论家因为很好的理由忽略了这个超膜理论。他们不知道是否能使这个理论与量子力学相容。有人尝试将它与量子论结合，失败了。当第一次超弦革命在1984年发生时，基于10维理论的迷人性质，多数理论家抛弃了这些11维理论。

可是现在，弦理论家在威藤的带领下，想起要复活11维的膜理论。他们这么做是因为他们发现了几个令人惊奇的事实。首先，如果我们将11维之一看作圆，就能将膜的一个维绕在那个圆上（图9-2）。这样，膜的另一维可以在其余9个空间维里自由运动。那就成了在9维空间里运动的1维物体。它看起来就是一根弦！

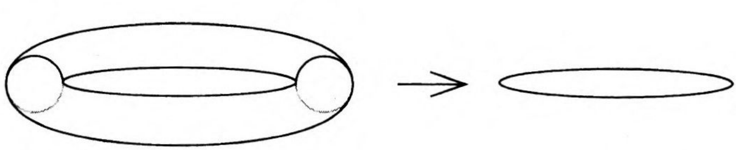


图9-2　左边是一个2维膜，我们可以想象它绕在一个隐藏维的小圆上。从远距离看（右图），

它就像缠绕在大空间维的一根弦

威藤发现，将膜的一个维以不同方式缠绕在圆上，就可以得到那5个和谐的超弦理论；而且只能得到那5个理论，没有别的。

事情还不仅如此。我们说过，当弦缠绕圆时，有一种叫T对偶的变换。这些对偶和其他的对偶不同，它们是精确的。当膜的一维缠绕圆时，我们也看到了这种对偶变换。如果用从缠绕的膜得到的弦理论来解释这些变换，它们恰好就是联系那些弦理论的强弱对偶。你大概还记得，那种特别的对偶原来除了某些特殊情形外都是猜想的，还没有证明。而现在我们认识了它们来自11维理论的变换。这个美妙的结果令人不得不相信11维统一理论的存在。我们剩下的唯一问题就是怎么去把它找出来。

那年下半年，威藤为那个未知的理论起了一个名字。起名字是很了不起的艺术：他干脆就称它为M理论。他不想解释M代表什么，因为理论还没有呢。我们的责任是构建那个理论来填补那个名字的空白。

威藤的讲话提出了很多问题。假如他是对的，就有很多东西等着我们去发现。听讲的人中有个叫波尔琴斯基（Joseph Polchin-ski），是圣塔巴巴拉的弦理论家。他告诉我们，“爱迪讲话过后，为更好理解它，我为自己列出了20个家庭作业问题。”[55]那些作业领着他做出了一个对第二次超弦革命起着关键作用大发现——弦理论并不仅仅是关于弦的理论，10维时空里还有别的东西存在。

不熟悉水族馆的人认为那儿的东西都是鱼。但水族爱好者知道，鱼不过是吸引你第一眼的东西。健康的水族馆都养着植物。如果你只养鱼，它是不会好起来的，很快就会成为一个死鱼塘。现在看来，在第一次超弦革命期间，从1984年到1995年，我们就像只知道养鱼的业余爱好者，遗漏了很多系统必需的东西，直到波尔琴斯基才发现了那些遗失的要素。

1995年秋，波尔琴斯基证明，弦理论为了和谐一致，不但需要弦，还需要在背景空间中运动的高维曲面。[56]这些曲面也是动力学对象。它们和弦一样，在空间自由运动。如果说一维的弦是基本的，那么二维的曲面为什么不能也是基本的呢？在高维下，空间更大了，为什么不能有三维、四维、甚至五维的曲面呢？波尔琴斯基发现，除非有了这些高维的东西，否则弦理论间的对偶性不可能和谐出现。他称那些东西为D膜。（膜是二维的东西，D代表的专业意思我不想在这儿解释。）膜在弦的历程中起着重要作用：它们是开弦的端点所在的地方。通常情况下，开弦的端点在空间自由运动，但有时弦的端点可能被约束在膜的曲面上（图9-3）。这是因为膜可以携带电荷和磁荷。

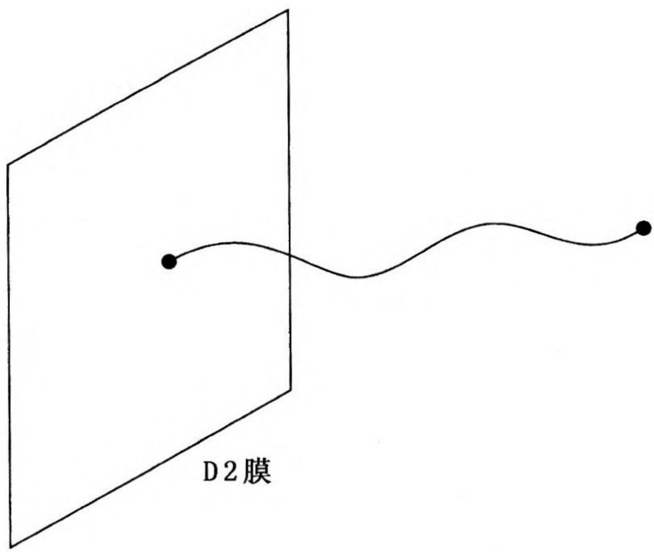



图9-3　一个二维膜上约束着一根弦的端点

从弦的观点看，膜是背景几何的额外特征。它们的存在带来了更多可能的背景几何，从而极大丰富了弦理论。我们除了可以缠绕某个复杂几何中的额外的维，还可以将膜缠绕在那个几何的圆和曲面上。你想要多少膜就有多少膜，它们可以任意多圈地缠绕紧化的维。这

样，我们就为弦理论制造了无限多个背景几何。波尔琴斯基的这幅图景将产生巨大的影响。

膜也深化了我们对规范理论与弦理论之间的关系的认识。事情是这样的：几张膜叠加在一起，使弦理论以一种新的方式产生对称。正如我刚才讲的，开弦可以终结在膜上。但如果几张膜在同一个位置，那么弦终结在哪张膜就无关紧要了。这意味着某种对称性在起作用，而我们在第四章说过，对称性可以产生规范理论。于是，我们就这样发现了弦理论与规范理论的新联系。但膜的作用还不止这一点。例如，我们将在下一章看到，它们还会引出更多可能的统一理论。

膜还为我们开辟了一条崭新的思路，去思考我们的三维世界如何能与弦理论的多个空间维相联系。波尔琴斯基发现的有些膜是三维的。将三维膜叠在一起，可以得到一个漂浮在多维空间里的具有任意对称性的三维世界。我们的三维宇宙会不会就是高维空间里的那样一个曲面呢？这是一个伟大的思想，有可能关联着一个叫膜世界的研究领域，在那儿，我们的宇宙就被看作漂浮在高维空间里的一个曲面。

膜实现了这一切；膜还做了更多的事情。因为它们，才可能在弦理论中描述某些特殊的黑洞。这是斯特罗明戈和维法（Cum-rum Vafa）在1996年发现的，那也许是第二次超弦革命的最大成就。

膜与黑洞的关系是间接的，然而也是牢固的。事情是这样的：我们先将引力去掉（可以令弦耦合常数为零）。以这种方式描述黑洞似乎很荒唐，因为黑洞什么也没有，惟独只有引力。不过我们接着往下看。没有了引力，我们可以考虑膜缠绕额外维的几何。现在我们利用膜携带电荷和磁荷的事实。结果发现，膜带多少荷是有极限的，与膜的质量有关。具有最大可能荷的构形很特别，我们称它为极端构形。它们构成我们以前讲过的一种特殊情形，存在能使我们进行更精确计算的额外的对称性。特别是，那些情形的特征在于具有联系费米子和玻色子的几种不同的超对称性。

黑洞能携带多少电荷或磁荷而依然保持稳定，也存在一个极大上限。那样的黑洞叫极端黑洞，广义相对论专家已经研究过多年了。如果研究在这些背景下运的粒子，你也会发现几种不同的超对称性。

令人惊奇的是，尽管引力消除了，极端膜系统却保留着极端黑洞的某些性质。特别是，两个系统的热力学性质是相同的。于是，通过研究缠绕额外维的极端膜的热力学，我们可以重现极端黑洞的热力学性质。

黑洞物理学的一个挑战是解释贝肯斯坦和霍金关于黑洞熵和温度的发现（见第六章）。根据弦论的新观点——至少在极端黑洞的情形——通过研究类似的缠绕额外维的极端膜系统，我们能取得进步。实际上，两个系统的许多性质都可以对应起来。之所以出现这种几乎奇迹般的巧合，是因为两种情形都存在几种不同的联系费米子和玻色子的超对称变换。结果，我们可以构造强有力的数学类比，迫使两个系统的热力学完全等价。

但事情不仅如此。我们还可以研究几乎极端的黑洞，即它们携带的荷比最大可能的数略小。对膜来说，我们也可以研究具有比最大荷略小的膜的集合。那么膜与黑洞的对应还存在吗？答案是肯定的，而且确实存在。只要离极端情形很近，两个系统的性质也几乎可以对等。这是对应的更严格验证。不论在哪个系统，温度与其他量（如能量、熵和荷）之间都存在复杂而精确的关系。两种情形非常一致。

1996年，我听了年轻的阿根廷博士后马尔德希纳（Juan Maldacena）就这些结果发表的演讲，那是在意大利的里雅斯特（我常在那儿避暑）的一次会议上。我被征服了。膜的行为与黑洞的物理学在那么高的精度上对应，立刻令我心动，决定挤出一些时间重新回到弦理论上来。我请马尔德希纳共进晚餐，来到一家俯瞰亚德里亚海的比萨店。我发现他是我遇到的最聪明、最敏锐的年轻弦理论家之一。那天晚上，我们喝着酒，吃着比萨饼，讨论的一个问题是，膜系统应该

不仅仅只是黑洞模型吧？它们是不是为黑洞的熵和温度提供了真正的解释呢？

我们不能回答那个问题，它现在仍然没有答案。答案要看那些结果到底有多重要。我们在这儿遇到的情形，我在其他场合已经说过了，即额外的对称性引出重大的发现。这里还是存在两种观点。悲观的观点认为两个系统的关系可能是它们同样具有很多额外对称性的偶然结果。对悲观者来说，计算的优美并不意味着它们带来了黑洞的一般认识。相反，悲观者担心，计算之所以优美是因为它们依赖于非常特殊的条件，而那些条件不能推广到典型的黑洞。

然而，乐观者相信，所有黑洞都可以用同样的思想来理解，特殊情形下表现的额外对称性只不过使我们把计算做得更精确。和强弱对偶的情况一样，我们还是不能确定悲观者和乐观者谁对谁错。这里，我们还有一点忧虑，即膜的叠加不是黑洞，因为引力已经被清除了。人们猜想，当引力慢慢恢复时，那些膜可以变成黑洞。实际上，可以想象这种事情会在弦论中发生，因为引力的强度正比于某个在空间和时间中变化的场。但问题在于这样的过程——其中引力场随时间变化——总是很难用弦理论进行具体描述。

虽然马尔德希纳的黑洞研究很精彩，那才是他的开始。1997年秋，他发表了一篇惊人的论文，提出了一类新的对偶性。[57]我们前面讲的那些对偶性都发生在同类理论之间，处于相同维数的时空。马尔德希纳的革命性思想是，弦理论可以有一个规范理论的对偶描述。这是令人惊讶的，因为弦理论是引力的理论，而规范理论却在固定背景的时空里描述没有引力的世界。而且，弦理论描述的世界比代表它的规范理论有着更多的空间维。

为了理解马尔德希纳的建议，我们回想一下第七章讨论过的思想，其中，弦理论可以从电场的流线产生出来。在那儿，电场的流线成了理论的基本对象。因为流线是一维的，看起来就像弦。你可以说

线变成了突现的弦。在多数情况下，来自规范理论的突现弦并不像弦理论讨论的弦。特别是，它们似乎与引力毫无关系，而且没有提供力的统一。

然而，波利亚柯夫早就提出，在某些情形，伴随规范理论的突现弦可能像基本弦。但规范理论弦并不存在于我们的世界；相反，波利亚柯夫凭着弦论历史上最惊人的想象力，猜想那些弦可能会在高一维的空间中运动。[58]

波利亚柯夫是怎么成功猜想到他的弦要在多一维的空间里运动呢？他发现，如果用量子力学方法处理从规范理论生成的弦，则它们具有一种突现性质，而那种性质竟然可以用弦上每一点的一个数字来描述。那个数字还可以理解为距离。在这种情形，波利亚柯夫提出，弦的每一点所赋予的数字，应该认为确定了那一点在额外维的位置。

考虑这种突现性质后，就会很自然地将电场的流线看成是高一维空间里的东西。于是，波利亚柯夫顺着这个思路提出了三个空间维的世界里的规范场与四个空间维的世界里的弦理论之间的对偶性。

如果说波利亚柯夫提出了一般性的思想，那么马尔德希纳将那思想具体化了。在他研究的世界里，我们的三个空间维包容着最大超理论——即具有最多超对称性的规范理论。他研究了可能成为规范理论的对偶描述的突现弦。通过推广波利亚柯夫的论证，他发现描述那些突现弦的弦理论实际上就是一个10维超对称弦理论。在弦所在的9个空间维中，有4个就像波利亚柯夫猜想的样子。这样就还剩下5个空间维，就是卡鲁扎和克莱因描述的额外维（见第三章）。将这些额外的维设计成一个球面，这样的空间有时被称作鞍形空间（图9-4）。它们对应于具有暗能量的宇宙，但那儿的暗能量是负的。

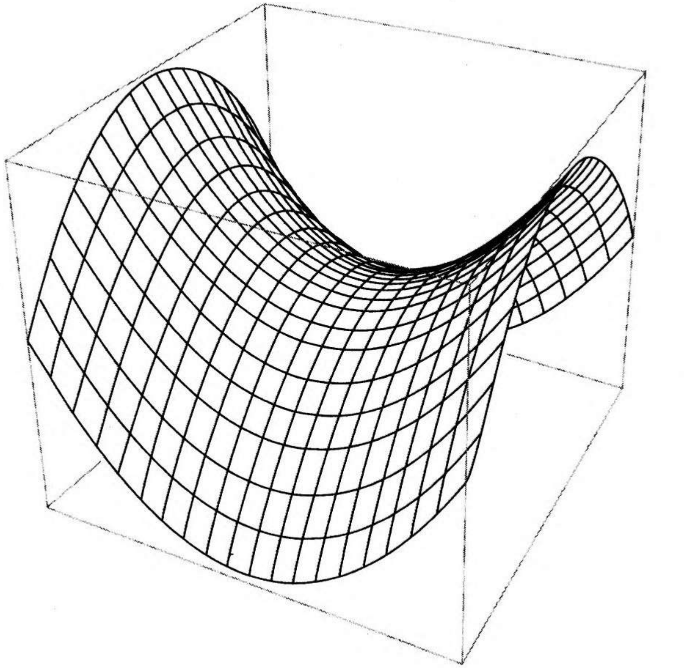


图9-4　鞍形曲面，代表具有负能量密度的宇宙的空间几何

马尔德希纳的猜想比波利亚柯夫最初的设想要大胆得多，它激起了巨大的反响，也成了后来千百篇论文的主题。虽然猜想至今尚未证明，但大量证据表明，在弦理论与规范理论之间至少存在一种近似的对应。

这里有着很大的风险——过去有，现在也有。如果马尔德希纳的对偶猜想是正确的，两个理论是等价的，那么我们将有一个量子弦理论的精确的量子描述。我们想问的关于超弦理论的任何问题都可以转换为一个关于最大超理论——它是规范理论——的问题。从原则上讲，这比我们在其他情形得到的东西要多得多——在那些情形，弦理论只不过是通过一系列的近似在背景相关的水平上定义的。

然而，还有几点警告。即使那种对偶性是真的，猜想也只有在对

偶的一方能精确定义时才有用。要为在10维空间运动的弦定义一个和谐的理论，却面临着巨大的障碍。于是，人们将希望寄托在相反的方向，根据猜想，用最大超理论来定义弦理论。可是，虽然我们对最大超理论知道很多，那个理论却也还没精确定义。看来，我们有希望做得更好，但还存在很多技术难题。

如果马尔德希纳的猜想错了，那么最大超理论和弦理论不会等价。然而，即使如此，也有显著的证据表明，两者在某些近似的水平上存在有用的关联。这些近似也许还不足以用一个理论来定义另一个，但它们仍然有可能计算一些喜欢关联的性质。人们沿着这条路线做了大量富有成果的工作。

例如，在最低的近似水平上，10维理论只是广义相对论在10维的一种拥有超对称性的推广形式。它没有量子力学，但有很好的定义。很容易在这个理论中做一些计算，如研究10维时空几何中不同类型的波的传播。值得注意的是，即使马尔德希纳的猜想只在最低近似下成立，也能使我们计算某些与我们三维世界的规范理论相对应的性质。

反过来，这也使我们更好地认识了其他规范理论的物理学。其结果是，至少在最低近似水平上，有很好的证据表明弦理论和规范理论确实像马尔德希纳猜想的那样相互关联。不论马尔德希纳猜想的强形式是对或错——哪怕弦理论本身错了——我们都为理解超对称规范理论找到了强有力的工具。

经过了几年的紧张工作，这些问题仍然混乱不清。关键在于弦理论与最大超理论之间的关系到底是什么。多数证据是用弱形式的马尔德希纳猜想解释的，它只要求一个理论的某些量可以用另一个理论的方法在一定的近似水平上进行计算。正如我说过的，这已经算是很好的结果了，有着重要的应用。但多数弦理论家相信强形式的猜想，即两个理论是等价的。

这种状况令人想起强弱对偶猜想，因为它使我们有可能在一个非常特殊的、具有很多额外对称性的状态子空间上说明强结果。与强弱对偶的情形一样，悲观者担心额外对称性会强迫理论以不容选择的方式达到一致，而乐观者相信额外对称性能帮助我们揭示在更一般情形下也正确的关系。

最后，究竟哪种形式的马尔德希纳猜想正确，有着多方面的影响。一个方面在于黑洞的描述。黑洞可以在具有负暗能量的宇宙中产生，所以我们可以用马尔德希纳猜想来研究如何解决霍金提出的黑洞信息疑难。两个理论是精确对应还是近似对应，将给疑难带来不同的解决方法。

假定黑洞内部的引力理论与规范理论只有部分的对应，那么在这种情形下，正如有些理论家（如惠勒和德维特）在很久以前设想的，黑洞可以永久保存信息——甚至将信息传递到从黑洞中心奇点生成的新宇宙。这样，信息总算没有丢失，因为它存在于新的宇宙；但对在黑洞边界的观察者来说，信息永远地丢失了。如果边界处的规范理论只包含内部的部分信息，那么这种信息丢失是可能发生的。但是，如果假定两个理论的对应是精确的，那么规范理论既无视界也无奇点，信息也就无处可丢。如果它正好对应于黑洞的时空，也就没有信息能在那儿丢失。在第一种情形，观察者失去了所有信息；在第二种情形，他保住了信息。直到我写这些话时，问题还没解决。

我们已经不止一次地看到，超对称在弦理论中扮演着一个基本角色。没有超对称性而构建的弦理论是不稳定的；不仅如此，它们还会通过一个永不停歇的过程发射越来越多的快子，直到理论崩溃。这可一点儿也不像我们的世界。超对称性清除了这些行为，使理论稳定下来。但在某些方面它又做得太过了。因为超对称性意味着有一种时间的对称，结果，超对称理论就不能建立在随时间演化的时空里。于是，使理论稳定的方面却使我们很难研究我们最希望用引力的量子理论回答的问题，如大爆炸后宇宙发生了什么？黑洞视界内部的深处发

生了什么？在这些情形，几何都是在时间中瞬息变化的。

这就是我们在第二次超弦革命中认识的弦理论。随着一系列前所未有的激动人心的结果，我们的认识极大地扩展了。它们为我们带来了一些诱人的真理的线索——如果能透过曾经的帷幕看到真实的东西，那该是多好啊！可是，尽管我们努力了，我们想做的很多计算还是可望而不可即。为了得到一个结果，我们不得不选择特殊的例子和条件。在很多情形，我们甚至不知道我们能做的计算是否能真的指引我们通向一般的情形。

我个人觉得这种状况是非常令人沮丧的。也许我们在向着一个万物的理论大踏步前进，也许我们盲目地夸大了结果，过分乐观地解读了我们能做的计算，在迷失的方向上越走越远。90年代中期，当我向弦理论的一些领头人物抱怨时，他们叫我别担心，说那只是因为理论比我们聪明。他们告诉我，我们不能直接向理论提问并要它回答。任何想直接解决大问题的努力是注定要失败的。相反，我们应该相信理论，跟着它走，心安理得地用我们不完美的计算方法去探索它愿意向我们袒露的部分。

只有一个陷阱。真正的量子形式的M理论应该是背景独立的，同样，任何引力的量子理论也必须是背景独立的。但除了我前面说的那些原因而外，M理论之所以必须是背景独立的，还因为5个超弦理论连同它们的所有流形和几何，都应该是M理论的组成部分。这包括那些几何在从1维到10维的空间里的所有不同的卷曲方式。它们都为弦和膜的运动提供背景。但如果它们是一个统一理论的部分，那个理论就不能建立在任何一个背景上，因为它必须囊括所有背景。

于是，M理论的关键问题就在于寻找一种能与量子理论和背景独立性相容的形式。这是很重要的问题，也许还是弦理论未解问题中最重要的一个。遗憾的是，这方面几乎没有什么进展。有一些迷人的线索，但我们还不知道M理论是什么，甚至不知道是否存在任何配得上

那个名字的理论。

量子力学的M理论有一定的进步，但仍然限于特殊的背景。早在80年代，这就是构造11维膜理论的量子理论的一种尝试。三个欧洲物理学家德威特（Bernad de Wit）、霍普（Jens Hoppe）和尼科莱（Hermann Nicolai）发现，可以通过一种技巧来做到这一点，即将膜表示为数学家所谓的矩阵——即一个二维的数表或数组。他们的公式需要9个这样的数表，由此得到一个逼近膜行为的理论。[59]

德威特和他的同事们发现，可以使他们的矩阵理论与量子理论相容。唯一的麻烦在于，为了描述膜，必须将矩阵扩张到无限，而量子理论只有在矩阵有限的时候才会有意义。所以，我们留下一个猜想：假如量子理论能和谐地向无限的数组扩张，那么它将生成一个膜的量子理论。

1996年，四个美国弦理论家复活了这个思想，不过有点儿费解。邦克斯（Thomas Banks）、费歇尔（Willy Fischler）、申克尔（Stephen Shenker）和苏斯金提出，在11维平直时空背景下，同样的矩阵理论不仅给出了11维膜理论，还给出了整个M理论。[60]这个矩阵理论并没有完全回答M理论是什么，因为它处于特殊背景之下。它只能在其他几种背景下有效，但当四个以上的空间维卷曲时，它就给不出合理的答案。如果M理论是对的，我们的世界应该有7个卷曲的空间维，所以它还不够好。而且，我们还不知道它在矩阵变得无限的时候，是不是能导出一个和谐的量子理论。

遗憾的是，M理论仍然只是一个诱人的猜想。它吸引着人们相信它；同时，在它真正建立起来之前，它的确不是一个理论——而是一个关于我们向往的理论的猜想。

当我在那些年思考和弦理论的关系时，想起我的一个艺术收藏家朋友。我们见面时，他说他还是我崇拜的一个年轻作家的朋友；我们

可以叫她“M”。几个星期后，朋友来电话说，“我几天前和M谈过话，你知道，她对科学非常感兴趣。你们什么时候见见面怎么样？”我当然有点儿受宠若惊，高兴地答应了。丰盛的晚餐刚吃到一半，收藏家的手机响了。“是M的，”他说。“她就在附近，想顺便过来看看你。行吗？”可她没来。饭后吃点心时，我和收藏家朋友畅谈艺术与科学的关系。后来，我不禁暴露了对M的好奇，真渴望见见她，样子很窘迫，于是告辞回家了。

几个星期以后，他打来电话，说了一大堆抱歉的话，然后又请我去吃饭见她。我当然去了。首先是因为他只在最好的饭店吃饭；看来这些画廊的经纪人的吃喝钱比科学家的薪水还高。不过，那次和后来几次，也都发生了同样的事情。她会来电话，一两个小时或几个小时过后，他的电话又会响起。“哦，我知道了。你感觉不舒服。”或者，“出租司机不知道剧场在哪儿吗？把你带到布鲁克林了？这城市怎么了？好的，我知道，很快……”两年后，我才相信那个年轻女子在书的封面上的照片是假的。一天晚上，我告诉他我终于明白了：他就是M。他只是笑了笑，说，“哦，是的……不过，她真的很高兴见你。”

弦理论的故事也许会像我和M的约会一样，永远拖延下去。虽然你知道那不是真的，却仍然为它工作，因为它距离你所能达到的目标最近。同时，伙伴很迷人，吃的东西也好。你一次次听说真正的理论就要揭开了，但从来没有发生。过后，你开始自己去找它。这样的感觉很好，可永远没有结果。最后，你得到的还是从前的那样东西：一本你不可能打开的书的封面上的一张美丽的图片。

# 第十章　万物之理

在两次弦的革命中，观察几乎不起任何作用。随着众多弦理论的成长，多数弦理论家仍然一如既往地相信他们原来的幻想：有一个能

带来唯一实验预言的唯一理论，但没有如愿的结果；而少数理论家则一直担心唯一的理论永远不会出现。同时，乐观者坚持认为我们必须相信理论，跟着理论走下去。一个统一理论需要做的事情，弦论似乎已经做得够多了，仿佛其余的东西到时候会自己冒出来。

然而在最近几年里，弦理论家的思想方法完全颠倒了。长久以来抱有的对唯一理论的希望破灭了，他们多数人现在相信弦理论应该看作一派广袤的理论景观，不同的理论描绘着一个多重世界的不同景象。

他们的期待因为什么而转变呢？是与事实的矛盾。但那并不是我们希望的事实——而是出乎我们多数人意料的事实。

好的理论应该令我们惊奇；它意味着不论谁创立了它，都算有了进步。但是，当观察令我们惊奇时，理论家就担心了。过去30年里，最令人困惑的观察莫过于1998年的暗能量的发现。我们说能量是暗的，意思是它看起来不同于我们以前知道的任何形式的能量和物质，即它不与任何粒子或波相联系。它就是它。

我们不知道暗能量是什么；我们知道它只是因为我们能观测它对宇宙膨胀的影响。它就像一个均匀遍布在整个空间的引力源。因为分布均匀，处处相同，所以没有东西向它落下。它唯一的效应是对星系分离的平均速度的影响。1998年，人们对遥远星系的超新星观测表明宇宙膨胀在加速，而其加速方式最好用暗能量的存在来解释。[61]

暗能量也许是所谓的宇宙学常数。这个名词指的是具有某种显著特征的能量形式：这种能量的性质（诸如密度）在所有观测者看来都是一样的，而不论他们在时间和空间的什么地方，也不论他们是否在运动。这是很不寻常的。通常情况下，能量伴随着物质，而且存在一个与物质一起运动的特殊观察者。宇宙学常数就不同了。它之所以叫常数，是因为不论你什么时候在什么地方以什么方式观测，它的值都

是一样的。又因为不能从空间运动的粒子或波寻求它的起源和解释，所以叫宇宙的——就是说，它是整个宇宙的特征，而不是其中任何具体事物的特征。（我要说明，我们还不能确定暗能量是否真的具有宇宙学常数的形式；虽然我们目前的所有证据都指向它，但在未来的几年，我们将更好地认识它的能量密度在时空中是否真的保持不变。）

弦理论没有预言暗能量；更坏的是，弦理论很难调和探测的数值。于是，暗能量的发现预示着弦论领域的危机。为看清这一点，我们必须回过头来讲讲宇宙学常数的奇异而悲凉的故事。

故事要从1916年说起，起因是爱因斯坦拒绝相信他新创的广义相对论的最大胆预言。他原本就倡导了广义相对论的一大理念，即时间和空间的几何是动力学演化的。所以，当人们用他的新理论去摹写宇宙，发现那宇宙也随时间而动力学演化，他是不应该感到惊奇的。人们研究的模型宇宙膨胀也收缩，甚至有起点和终点。

但爱因斯坦还是被结果惊呆了——也惊慌了。自亚里士多德以来，宇宙总被认为是静态的。它也许是上帝创造的，但如果真是那样，那它就从未改变过。爱因斯坦是200年来最富创造力也最为成功的理论物理学家，可就是这样一个人物，也不能想象宇宙竟然不是永恒不变的。我们不禁要说，假如爱因斯坦真是天才，它大概会更相信他的理论而不是偏见，从而预言宇宙的膨胀。而我们得到的更深刻的教训是，即使最冒险的思想家，要让他们抛弃延续了几千年的信仰，也是多么艰难啊！

我们这些已经习惯了宇宙演化思想的人，对于人们接受宇宙可能有起点的观点到底有多难，也只能在心里揣度了。不管怎么说，在相当长的一段时间里，并没有宇宙随时间变化或演进的证据，所以爱因斯坦把宇宙膨胀的预言看成他的理论存在缺陷的信号，并开始想办法让它与永恒宇宙的概念相容。

他发现，他的引力方程还有一种可能性，即真空的能量密度可能不等于零。而且，这个普适的能量密度对所有观察者来说都是一样的，而不管他们什么时候在什么地方进行观察，也不管他们在如何运动。所以他称它为宇宙学常数。他发现，常数的效应依赖于它的符号。当它是正数时，将引起宇宙膨胀——不仅膨胀，而且加速地膨胀。这不同于普通物质的影响，它们总是使宇宙收缩，因为它们总是相互吸引的。于是，爱因斯坦意识到，他可以利用这种新的膨胀趋势来平衡引力产生的收缩，从而达到一个静态而永恒的宇宙。

爱因斯坦后来说宇宙学常数是他一生最大的失误。实际上，那是一个双重的失误。首先，它没有什么成效，并没有阻止宇宙的收缩。你可以平衡物质引起的收缩与宇宙学常数引起的膨胀，但那只是暂时的。平衡根本就是不稳定的。稍微扰动一下，宇宙就会开始收缩或膨胀。不过真正的失误在于静态宇宙的思想一开始就是错的。10年后，一个叫哈勃（Edwin Hubble）的天文学家开始寻找宇宙膨胀的证据。自20世纪20年代以来，宇宙学常数就成了人们想摆脱的烫手的山芋。可是，随着时间的流逝，它变得越来越难摆脱了，至少在理论上是那样。我们不能只是规定它为零，然后忽略它。它就像角落里的大象，你可以视而不见，但它确实在那儿。

人们不久就开始认识到量子理论能对宇宙学常数有所解释。遗憾的是，它和我们想听的话正好相反。量子理论，特别是不确定性原理，似乎需要一个巨大的宇宙学常数。如果某个东西是完全静止的，那么它有确定的位置和动量，而这跟不确定性原理相矛盾，因为原理告诉我们，不可能同时知道粒子的这两样东西。结果，即使温度为零，事物也在不停地运动。任何粒子和任何自由度，即使在温度等于零的时候，也都伴随着一定的剩余能量。这叫真空能量或基态能量。当量子力学用于场（如电磁场）时，场振动的每个模式都有真空能量。但场振动的模式很多，所以量子力学预言的真空能量很大。在爱因斯坦广义相对论的背景下，这隐含着巨大的宇宙学常数。我们知道

这是错误的，因为那意味着宇宙膨胀很快，从而不可能形成任何结构。星系的存在为宇宙学常数的大小强加了一些限制，那些极限大约比量子力学预言的小120个数量级。这大概是科学理论最拙劣的一个预言了。

出现大问题了。理智的人会认为我们还需要崭新的思想，在引力与量子理论的偏离得到解释之前，两者的统一是不会有什么进步的。几个最敏锐的人也有同感。其中一个是德国理论物理学家德雷耶（Olaf Dreyer），他猜想，只有当我们抛弃空间是基本的观念，才可能消解量子理论与广义相对论的不相容。他提出，空间本身是从某种迥然不同的更基本的描述中涌现出来的。几个在凝聚态物理学领域做过重要工作的理论家，如诺贝尔奖得主劳克林（Robert Laughlin）和俄罗斯物理学家沃洛维克（Grigori Volowik），也讨论过这个观点。但我们大多数做基础物理学工作的人都不在乎这个问题，而是继续走我们不同的路线，哪怕到头来一事无成。

直到不久以前，我们还一直有一点安慰：至少宇宙学常数的观测值为零——就是说，没有宇宙加速膨胀的证据。这令人满意，因为我们可以指望找一个新原理来清除方程中的所有令人不安的东西，并使宇宙学常数严格等于零。假如观测值是某个非零的小数，问题就严峻多了，因为要一个新原理将数值减小但不等于零，是更难想象的事情。于是，几十年来，我们都对各路神灵满怀感激，至少我们没有遭遇那个问题。

宇宙学常数给所有物理学家提出了一个问题，不过弦论的情况似乎更好一点儿。弦论不能解释为什么宇宙学常数等于零，但它至少解释了它为什么不是正数。我们能从弦论得到的仅有的几个结论之一就是宇宙学常数只能为零或负数。我不知道有哪个弦理论家预言过宇宙学常数不能是正数，但一般都认为那就是弦论的一个结论。其中的道理太专业了，不能在这儿评说。

其实，负宇宙学常数的弦理论早就研究过了。例如，著名的马尔德希纳猜想就涉及一个具有负宇宙学常数的时空。困难还很多，直到今天也没人具体写出弦理论在负宇宙学常数世界里的细节。但人们相信，没写出具体的东西，只是一个技术问题——而不是原则问题。还没有什么理由说它在原则上就应该是不可能的。

于是你可以想象，当1998年的超新星观测表明宇宙在膨胀加速，意味着宇宙学常数必须是正数的时候，人们是多么惊奇。这是真正的危机，因为弦理论的预言与观测之间显然出现了矛盾。实际上，有定理表明，具有正宇宙学常数的宇宙——至少在忽略量子效应时——不可能是弦理论的解。

威藤不是悲观主义者，但他在2001年还是坦率承认“我不知道有什么好办法能从弦理论或M理论得到德西特空间[具有正宇宙学常数的宇宙]。”[62]

科学哲学家和科学史家，如拉卡托斯（Imre Lakatos）、费耶阿本德（Paul Feyerabend）和库恩，曾经指出，一个实验反常不足以否定一个理论。如果一个理论深得众多专家的信任，他们会用更极端的方式来挽救它。这对科学来说并非总是坏事，甚至还可能是好事。有时，理论的捍卫者会成功，而他们的成功往往伴随着意外的重大发现。但有时他们也会失败，大量的时间和精力都在科学家们的一步步努力中浪费了。最近几年发生的弦论的故事，应验了拉卡托斯和费耶阿本德的观点，因为它表现的是一大群专家在尽力挽救一个他们珍爱的然而却面对着矛盾数据的理论。

如果说弦论真的得救了，那么拯救它的是一个完全不同的问题的解：如何使高维稳定？回想一下，在高维理论中，额外维的卷曲产生了很多解。那些能再现我们观察的世界的解是很特殊的，其高维空间几何的某些方面必须冻结起来。否则，一旦几何开始演化，就可能一直继续下去，要么生成奇点，要么迅速膨胀，张开卷曲的维，使我们

能看到它们。

弦理论家称它为模的稳定性问题。“模”是刻画额外维性质的常数的一般名字。这是弦理论必须解决的问题，但长期以来还不清楚应该怎么做。和其他情形一样，悲观者很忧虑，而乐观者相信我们迟早会找到答案的。

这回是乐观者说对了。进步从90年代开始，加州的几个理论家认识到问题的关键在于用膜来稳定高维。为明白这一点，我们必须认识问题的一个特征，即高维几何在为弦理论提供良好背景的同时可以连续变化。换句话说，你可以改变高维的形状或体积，并通过这种方式经历不同的弦理论空间。这意味着没有什么能阻止额外维的几何随时间演化。为避免这种演化，我们必须找一组弦理论，而不可能连续游移其间。为实现这一点，我们需要寻求那样的弦理论，其每一步变化都是离散的——就是说，你不能连续地从一个理论走到另一个，而只能突然地大踏步跳越。

波尔琴斯基告诉我们，弦论中确实有离散的东西：膜。回想一下，存在那样的弦背景，其中膜是缠绕在额外维曲面上的。膜以离散的形式出现。你可以有1，2，17或2040197个膜，却不能有1.003个膜。因为膜携带电荷与磁荷，这就生成电磁流的离散单位。

于是，90年代末，波尔琴斯基和一个叫波索（Raphael Bousso）的博士后一起，开始研究额外维环绕着大量电流的弦理论。他们得到一个理论，其中的一些参数不再连续变化了。

但你能以这种方式冻结所有的常数吗？这需要很多复杂的构造，但结果有一个额外的好处。它达成了具有正宇宙学常数的弦理论。

关键的突破发生在2003年，主人翁是一群来自斯坦福的科学家，包括卡洛什（Renata Kallosh，超引力和弦理论的先驱之一）、林德

（Andrei　Linde，宇宙暴胀的发现者之一）和两个优秀的年轻理论家卡齐鲁（Shamit　Kachru）和特里维迪（SandipTrivedi）。[63]即使从弦论的标准看，他们的工作也太复杂了；他们的斯坦福同事苏斯金说它是"用疱丁杀鸡的技艺"。但它的影响很大，因为它不但稳定了额外维，还协调了弦理论与暗能量的观测结果。

斯坦福小组的工作可以简单说明一下。他们的出发点是一种经过了认真研究的弦理论——在每一点带有6维小几何的平直的四维时空。他们选择的6维卷曲的几何是卡丘空间（见第八章）。我们已经看到，至少有十万个那样的空间，而我们只需要选一个其几何依赖于很多常数的典型代表。

接着，他们让大量电磁流（通量）卷曲在每一点的6维空间周围。因为只能卷曲离散单位的电磁流，这就可能能清除不稳定性。为了进一步稳定这种几何，还必须借助某些量子效应，虽然我们不知道它们如何直接从弦理论产生，但在超对称规范理论中对它们已经有了一定的认识，因此它们有可能起一定的作用。将这些效应与流和膜的效应结合起来，就能得到所有模都稳定的几何。

通过这一点，也可能在四维时空出现负的宇宙学常数。原来，我们希望的宇宙学常数越小，需要卷曲的流就越多，所以我们卷曲大量的流得到的宇宙学常数虽然小然而却是负的。（前面说过，我们还不清楚如何在这样的背景下具体写出弦理论，但也没有理由相信它不存在。）但问题的关键在于得到正宇宙学常数，以满足宇宙膨胀速率的新观测结果。于是，下一步就是以不同方式卷曲其他的膜，这样可以增大宇宙学常数的值。正如有反粒子，也有反膜，而斯坦福小组在这儿就用了反膜。通过卷曲反膜，可以增加能量，从而使宇宙学常数成为小的正数。同时也抑制了弦理论相互游移的趋势，因为任何改变都必须是离散的跳越。这样，一举解决了两个问题：清除了不稳定性，宇宙学常数成了正的小数。

斯坦福小组本可以将弦理论从宇宙学常数引发的危机中解救出来，至少暂时可以，但他们的方法产生的结果太奇怪，也太出人意料，反倒使弦论的阵营四分五裂。在此之前，弦理论家们是步调一致的。20世纪90年代参加弦理论会议就像80年代初去中国，几乎每一个和你谈话的人都抱有同样的观点。不管是好还是坏，斯坦福小组破坏了团结。

记住，我们现在讨论的特殊的弦理论，是从紧化空间周围的卷曲流产生的。为了得到小的宇宙学常数，必须卷曲很多流。但流的卷曲方式不止一种；实际上有很多种。那么到底有多少呢？

回答这个问题之前，我必须强调，我们并不知道通过卷曲在隐藏维周围的流所生成的理论中，是否有哪个能给出优美而和谐的量子弦理论。用我们现有的方法很难回答这个问题。所以，我们要做的事情是求助检验，它能说明一个好的弦理论的必要但不充分的条件。我们的检验要求弦理论（如果存在的话）具有相互作用很弱的弦。这意味着如果能在这些弦理论中进行计算，那么结果应该非常接近我们所能进行的近似计算的预言。

我们能回答的问题是，这样有着在6个隐藏维周围的卷曲流的弦理论中，有多少通过了考验？答案依赖于我们想要多大的宇宙学常数。如果我们想一个负的或零的宇宙学常数，那么就有无限多个不同的理论。如果我们希望理论有一个正的宇宙学常数以满足观测结果，那么理论的数量就是有限的；目前证据表明大概有10500个。

这当然是一个巨大的数目。而且，每个弦理论都是不同的。每一个都将为基本粒子物理学给出不同的预言，也都将为标准模型的参数值给出不同的预言。

我们说弦论不是一个理论，而是一个由众多可能理论构成的景观，这个观点在20世纪80年代末和90年代初就已经提出，但被多数

弦理论家拒绝了。前面说过，斯特罗明戈在1986年发现有许多看起来很和谐的弦理论，虽然很多弦理论家仍然相信会出现确定唯一一个正确理论的条件，还是有少数人一直担心会失去预言能力。但波索、波尔琴斯基和斯坦福小组的工作最终打破了平衡。他们也和斯特罗明戈一样，带来了大量新的弦理论，不过真正新奇的地方在于这些数量的需要是为了解决两个大问题：让弦理论与正真空能的观测结果一致；将理论固定下来。也许因为这些原因，理论的景观最后就不能看作是可以忽略的奇异结果，而是让弦理论免遭否决的补救方法。

景观论的另一个原因倒非常简单，那就是理论家们泄气了。他们费了很长时间寻找能替他们选出唯一弦理论的原理，但一直没有发现那样的原理。随着第二次革命的进行，我们现在对弦论的理解好多了。特别因为对偶性的出现，就更难说多数弦理论都是不稳定的。于是，弦理论家开始接受所有可能理论所形成的景观。驱动这个领域的问题不再是如何寻找唯一的理论，而是如何做有那么多理论的物理。

一种响应是说那是不可能的。即使我们只限于与观测一致的理论，似乎也还有太多那样的理论，总有几个几乎肯定能给出我们想要的结果。那么为什么不把这种情形看作reductio ad absur-dum（“归谬”）呢？那个词在拉丁文里很好听，不过在英文里更实在，还是直说吧：本来是想构造一个唯一的自然理论，努力的结果却引出了10500个理论，那方法岂不是归于荒谬了么？

这对那些把多年甚至几十年的生命都奉献给了弦理论的人来说，确实是很痛苦的。我也曾为它付出过一定时间和精力，如果说我也感到痛苦，那不过是想象我那些把整个生涯都寄托在弦论的朋友们会有什么样的感觉。不过，尽管令人痛苦不堪，对这样的局面，理智和诚实的态度还是应该承认它“归谬”了。这样的态度，只是我认识的少数几个人坚持的，而多数弦理论家没有选择它。

还有一种理性的态度：否认存在大量弦理论的说法。关于具有正

宇宙学常数的新理论的论证是基于强烈的近似，它们也许会使理论家们相信数学上并不存在的理论，更不必说什么物理了。

实际上，我们说存在大量具有正宇宙学常数的弦理论，其证据来自非常间接的论证。我们不知道如何真实地描述在那些背景下运动的弦。而且，我们可以规定某些弦理论存在的必要条件，但我们不知道这些条件对理论的存在是否也是充分的。于是，我们并没有证明在那些条件下确实存在某个弦理论。所以，理智的人大概会说，也许它们不存在。的确，最近，霍洛维茨（卡丘空间的发现者之一）和两个年轻同事赫尔托（Thomas Hertog）和梅达（Kengo Maeda）发现的结果，就引出了这样的问题：这些理论中是否有哪个描述了稳定的世界？[64]对这些证据，我们可以重视，也可以像许多弦理论家那样满不在乎。赫尔托和他的同事们发现的可能的稳定性不仅损害了斯坦福小组发现的新理论景观，而且破坏了涉及6维卡丘空间的所有解。如果这些解确实都不稳定，那么它意味着旨在联系弦理论与现实世界的多数工作都不得不被抛弃。当前人们也在争论斯坦福小组的某些假定是否真的成立。

第一次超弦革命开始的时候，弦理论的存在是一个奇迹。而最终存在5个弦理论的事实更令人惊奇。那几乎是不可能的，就更增强了我们对计划的信心。起初不大可能，而后来却成立了——那么它当然是奇迹。今天弦理论家乐意接受众多理论构成的景观，而20年前我们相信存在一个单一的理论，不过都一样没有什么证据。

所以，最后的界线就是，“我需要你用20年前估价那5个理论的标准来说服我相信那些理论是存在的。”如果你坚持这些标准，那么你就不会相信大量的新理论，因为当前景观的任何一个理论的证据，在旧标准下都是微不足道的。这是我本人多数时间倾向的观点。在我看来，这似乎是对证据的最理性的解读。

# 第十一章　人类的选择

我认识的很多物理学家降低了他们的希望，不再指望弦论是自然的基本理论——但不是所有的人都那样。近年流行的争论是，问题不在于弦论，而在于我们对物理学理论的期待。这个论题是苏斯金几年前在一篇题为“人类选择的弦论景观”中引出的：

根据大量学者最近的工作，那一派景观似乎广袤无垠而千姿百态。不管我们是不是喜欢，这种状况又为人存原理增加了几分信心……这些[斯坦福小组的景观中的理论]一点儿也不简单。它们是临时的庖丁杀鸡的技艺，几乎不可能有什么重要的意义。但在人择的理论中是不考虑简单和优美的。选择真空的唯一原则就是实用，即它是否具有生命必须的元素，如星系的形成和复杂的化学等。我们需要的就是这些东西外加一个那样的宇宙学——它保证至少有一大块空间能以很高的几率形成那个真空结构。65

苏斯金说的人存原理，是宇宙学家们自1970年以来提出并探讨的一个老思想。当时他们面对着这样一个事实：生命只能出现在所有可能物理学参数的一个极端狭小的范围内，尽管奇怪，我们还是出现了，仿佛宇宙就是为了包容我们而设计的（所以才说“人存”）。苏斯金的特殊说法代表着一种宇宙图景，即林德前些年提出的所谓“永恒暴胀”。根据这幅图景，宇宙早期的迅速暴胀阶段生成的不仅仅是一个宇宙，而是一个无限的宇宙群。你可以想象宇宙的原初状态处于没有终止的指数式扩张。接着，气泡出现了，在那些地方膨胀急剧慢了下来。我们的世界就是其中的一个气泡，而另外还有无限多个。在这个图景下，苏斯金增加了一点：当气泡形成时，某个自然律就从众多弦理论中选出一个来主宰那个宇宙。结果是一大群宇宙，每一个都由一个随机选出的弦理论负责。那个景观里的每个可能的理论都能在所谓多重宇宙中找到自己的地方。

苏斯金们鼓吹人存原理，我看是很不幸的，因为我们已经认识到

了它对科学来说是一个脆弱的基础。由于每个可能的理论都决定着多重宇宙的某个部分，我们就做不出多少预言了。原因很简单。

如果一个理论假定了一大群由随机选择的定律决定的宇宙，那么，为了做出预言，我们首先必须写出所有关于我们宇宙的事实。这些事实也同样适用于很多别的宇宙，我们把所有满足这些事实的宇宙的集合称作可能真实的宇宙。

我们只知道我们的宇宙是可能宇宙中的一个。如果说一群宇宙是通过基本自然定律的随机分布产生的，那么我们就不可能知道别的什么了。只有当每个（或几乎每个）可能真实的宇宙都有某个在我们自己的宇宙中没有发现过的性质，才可能做出新的预言。

例如，假定在几乎每个可能真实的宇宙中，最高共振频率是低音C，那么从所有可能真实的宇宙中随机选择出来的宇宙很可能以低音C的频率振动。因为我们除了知道自己的宇宙是一个可能真实的之外对它一无所知，所以我们很可能预言我们的宇宙也是唱低音C的。

问题是，理论在所有宇宙的分布是随机的，很少有这样的性质。最可能的是，一旦我们确定了我们在自己宇宙中观测的性质，那么任何宇宙可能具有的其余性质将随机分布在其他可能真实的宇宙中。于是我们无法做出预言。

我所讲的是宇宙学家们所谓的弱人存原理。正如名字所说，关于我们的宇宙，我们知道一件事情，那就是它支持智慧生命的存在；于是，每个可能真实的宇宙都必须是智慧生命能够生存的地方。苏斯金等人指出，这个原理一点儿也不新鲜。例如，我们如何解释我们处于这样一颗行星——其温度恰好使水处于液态？如果我们相信宇宙中只有一颗这样的行星，就会发现这是很令人疑惑的事情。我们不禁要相信一定存在一个智能的设计者。但一旦我们知道有大量的恒星和众多的行星，我们就会明白只是出于偶然，才会有很多适宜生命的行星。

那么，我们出现在其中一颗也就不足为怪了。

然而，行星类比和宇宙学情形还是有一点很大的区别，那就是除了自己的宇宙，我们不知道别的任何宇宙。存在一组宇宙的假设，是不能通过直接观察来证明的，因而不可能用来做任何解释。的确，假如真的存在一组有着随机定律的宇宙，我们就不会惊奇自己生活在我们能够生活的一个宇宙中。但我们生活在适宜于生命的宇宙这个事实，不能用来证明存在一组宇宙的理论。

还有一个相反的论证，我们可以用行星的例子来说明。假定不可能看到任何其他行星，那么，当我们根据这一点推断只存在一颗行星时，我们将不得不相信一件非常不可能的事情，即那颗存在的唯一行星是适于生命的。另一方面，如果假定存在很多有着随机性质的行星，即使我们不可能看到它们，其中多半会有几个适宜于生命——实际上，那几乎是肯定的。于是我们可以说，存在众多行星的可能性比只有一颗行星的可能性要大得多。

但这个看似强硬的论证却是错误的。66为看清这一点，我们将它与另一个根据相同证据进行的论证进行比较。相信智能设计的人会说，假如只有一颗行星而且适宜生命，那么很可能有一个智能设计者在发生作用。如果我们在两种情形下进行选择：①那颗唯一的行星只是因为运气才适宜生命，②有一个智能的设计者，他制造了那颗行星并使它适宜生命。那么，同样的逻辑使我们觉得第二个选择更有道理。

多宇宙图景与智能设计图景起着同样的逻辑作用。二者都提出了不能检验的假设，如果假设对了，它将使不可能的事情成为可能。

这些论证之所以错了，部分原因在于它们依赖于一个没有明说的假定——我们把握了所有可能的情形。回到行星的类比，我们不能排除未来某个时候也许会出现一个真正的解释，解释我们的行星为什么

适宜生命。两个论证的错误在于它们都拿一个可能（但未经检验）的解释来对比一个不可能有解释的命题。当然，在只有这两个选择的情况下，一个解释会比任何未经解释的不可能性显得更为合理。

几百年来，我们都有很好的理由相信存在大量行星，因为存在很多恒星——最近我们直接证明了存在太阳系外的行星。所以我们认同用多行星来解释我们行星的生命特征。但当问题涉及我们的宇宙为什么有生命时，我们至少有三种可能：

1.我们的宇宙是大量有着随机定律的宇宙的集合。

2.存在智能设计者。

3.存在我们迄今未知的机制，既能解释我们宇宙为什么适宜生命，也能做出可以证明它或证伪它的可检验的预言。

前两种可能在原则上是不牢固的，最合理的是主张第三种情形。实际上，那是我们作为科学家需要考虑的唯一可能，因为接受前面的任何一种都意味着我们学科的终结。

有的物理学家声称必须严肃对待弱人存原理，因为它在过去指引我们做出了真正的预言。我这儿说的是我最仰慕的一些物理学家——除了苏斯金，还有温伯格，你大概还记得，第四章说过他和萨拉姆一起统一了电磁力和弱核力。于是，我很痛苦地拿出我的结论：在我考察的所有情形，他们的主张都是错误的。

例如，我们根据英国天体物理学家霍伊尔（Fred Hoyle）在50年代的考察，考虑下面有关碳核性质的论证。人们通常拿这个论证来说明真正的物理学预言可以建立在人存原理的基础上。论证从一个观察事实开始：生命的存在需要碳。实际上，碳很丰富。我们知道它不可能从大爆炸产生出来，所以我们知道它必然是在恒星中制造的。霍伊尔发现，只有当存在碳核的某个共振态时，才能形成碳。他把预言告

诉了一群实验家，他们找到了。

霍伊尔预言的成功有时成了人存原理的支持。但前面一段从生命出发的论证与那段的其余部分没有逻辑关系。霍伊尔其实是从宇宙充满了碳的事实得出一个结论，而他的基础是假定必然存在某个碳的生成过程。我们和其他生命由碳组成的事实，并不是论证所必须的。

经常引用的另一个支持人存原理的例子是温伯格1987年在一篇著名论文里对宇宙学常数的预言。在那篇论文里，温伯格指出，宇宙学常数必然小于某个值，否则宇宙将急速膨胀下去，不可能形成星系。[67]因为我们看到的宇宙充满了星系，所以宇宙学常数必须小于那个值。确实如此，也必须如此。这是很漂亮的科学。但温伯格将它有效的科学论证推得更远。他说，假定存在一个多重宇宙，假定宇宙常数值随机分布在它们中间，那么在可能真实的宇宙中，宇宙常数的典型值将具有与星系形成相应的最大数量级。于是，假如多重宇宙的图景是对的，那么宇宙学常数就具有可能的最大值，而同时允许星系的形成。

温伯格发表这个预言时，人们还普遍相信宇宙学常数等于零。因此，当他的预言大约在10%的精度内正确时，确实令人惊讶。然而，当新的结果要求对他的结论进行更仔细的验证时，问题就来了。温伯格原来考虑的一组宇宙，只有宇宙学常数是随机分布的，而其他参数都是固定的。其实，他本该在容许星系形成的所有宇宙中平均，允许所有参数变化。如果这样进行预言，宇宙学常数值就将差得很远。

这说明了这类推理的一贯问题。如果我们的图景需要随机分布的参数，而我们只能观测其中的一组，那么我们可以得到很大范围的预言，就看我们对未知和不可见的那些参数做什么样的假定。例如，我们每个人都属于很多群体，在有些群体中我们也许是代表性成员，但在其他许多群体中我们是非代表性的。例如，在为本书写的作者小传中，我写的都是有代表性的东西。你能从中了解我多少呢？

还有许多其他情形，也许验证了某个弱形式的人存原理。在基本粒子物理学的标准模型中，有些常数如果真是通过在一组宇宙中的随机分布进行选择的，就完全不会有我们预期的数值。我们预期夸克和轻子的质量除了第一代而外应该是随机分布的，但我们发现了它们之间的关系。我们预期基本粒子的某些对称性会被强核力打破更多，我们预期质子衰变比现在实验所限制的更快。实际上，我不知道哪个成功预言是从具有随机的定律的多重宇宙推导出来的。

但第三种情形又如何呢？它是以可以检验的假设为基础来解释我们宇宙的生命特征的。1992年，我公开提出了这样的设想。为了从多重宇宙得到可以检验的预言，宇宙族一定不能是随机的。它必须具有复杂的结构，从而存在所有或多数宇宙都具有的与我们的存在无关的性质。这样我们才能预言我们的宇宙具有那些性质。

得到这种理论的一个方法是向生物学的自然选择方式学习。我构造这个图景是在80年代末，那时已经清楚弦理论将以大量不同形式出现。从进化论生物学家道金斯（Richard Dawkins）和马古利斯（Lynn Margulis）的书中，我了解到生物学家已经有基于可能的显型空间（他们称之为适应性景观）的进化模型。我借用这个思想和名词，设想了一幅宇宙从黑洞内部生成的图景。在《宇宙的一生》（1997）中，我仔细考察了整个思想的意义，这里就不多说了；我只想说，那个我所谓的宇宙自然选择的理论，做出了真正的预言。1992年，我发表了其中的两个，它们也就从此确立起来了，尽管可能被后来的许多观察证明是错误的。两个预言是，68没有质量大于1.6个太阳质量的中子星，和暴胀生成的——也许就是在宇宙微波背景中看到的——涨落谱，应该与可能的最简单形式的暴胀一致，只有一个参数和一个暴胀子场。

苏斯金、林德和其他研究者批评过宇宙自然选择的思想，因为他们声称永恒暴胀生成的宇宙数目将远远超过从黑洞生成的宇宙数目。为说明他们的反对，需要了解暴胀的预言有多大可靠性。有时候，这

种情形变成了没有永恒暴胀就很难有暴胀，而它的证据是，暴胀宇宙学的某些预言得到了证实。然而，从暴胀到永恒暴胀的转变，假定了在我们目前宇宙尺度成立的结论可以毫无阻碍地推广到更广大的尺度。这里存在两个问题：第一，从当前尺度向更大尺度外推，在某些暴胀模型中也隐含着向更早、更小的尺度外推。（我这里不做解释，但它对几个暴胀模型是对的。）这意味着为了得到一个比我们当前宇宙大得多的暴胀宇宙，我们必须将早期宇宙的描述推广到比普朗克尺度小得多的时间，而在那之前，量子引力效应主导着宇宙的演化。这是有疑问的，因为暴胀的通常描述假定时空是经典的，没有量子引力效应；而且，几个量子引力理论都预言不存在比普朗克时间更小的时间间隔。第二，有证据表明，暴胀的预言在我们当前能观测到的最大尺度上没有得到满足（见第13章）。于是，从暴胀外推到永恒暴胀，既有理论困难也有观测困难，所以它对宇宙自然选择并不构成强有力的反驳。

虽然人存原理没有、也不太可能引出任何真正的预言，苏斯金、温伯格和其他一流物理学家却很欢迎它，认为它不但标志着物理学的进步，也标志着我们对什么是物理学理论的认识的进步。温伯格在最近的一篇文章中说，科学史上的多数进步都是以关于自然的发现为标志的，但在某些转折点，我们发现的是科学本身……现在我们也许站在一个新的转折点，我们作为物理学理论的合法基础将发生剧烈的变化……弦景观展示的参数的可能值越多，它就越可能使人存原理成为物理学理论的新的合法基础：任何研究自然的科学家都必然生活在那景观的一角，那里的物理学参数的取值正好适合生命出现并进化为科学家。69

温伯格正是以标准模型的贡献闻名的，他的著作总是洋溢着动人而清醒的理性。不过，简单地说，当你也像这样推理时，你就不能让你的理论经受某些检验——而科学史反复证明，必须经过这些检验才可能区分哪些理论是正确的，哪些理论是优美却是错误的。为此，理

论必须做出能被证实或否定的具体而精确的预言。如果预言有很高的被否定的风险，那么证实的意义就十分重大了。如果既不会被证实，也不会被否定，那就没办法继续做科学了。

科学怎么会遭遇一片广袤的弦景观呢？这个问题的争论在我看来有三种可能：

1.弦理论是对的，随机多重宇宙也是对的，为了包容它们，我们必须改变科学研究的法则，因为根据通常的科学规范，如果理论没有做出可以证明或证伪它的独特预言，我们就不应该相信它。

2.我们最终会找到一个方法，从弦理论导出真正的可以检验的预言。为了实现这一点，我们可以通过证明确实存在一个独特的理论，或者通过一个不同的能产生真正可检验预言的非随机的多重宇宙理论。

3.弦理论不是自然的正确理论。自然也许最好是由别的某个尚未发现或接受的理论来描述，它能带来真正的最终能被实验证实的预言。

我觉得令人深思的是多数著名科学家似乎都不能容忍弦理论或多重宇宙的假没是错误的。下面是他们的一些言论：

“根据人存原理的推理与理论物理学的历史目标是那么背道而驰，即使在认识到它可能是必须的之后，我很长时间仍然反对它，不过现在好了。”

——波尔琴斯基

不喜欢人存原理的人简直就是睁眼瞎。

德

“一派广袤的风光可能出现在我们面前，这是理论物理学激动人心的进步，迫使我们重新思考我们的许多假设。我从内心感到它也可能是正确的。”

——阿卡尼-哈姆德（NimaArkani-Hamed，哈佛大学）

“我想那景观很可能是真的。”

——特格马克（Max Tegmark，麻省理工学院）

威藤似乎很困惑：“我没有什么深刻的话好说了。我希望我们能学会更多。”[70]

说这些话的人没有一个不是我所敬仰的。然而，在我看来，任何公正的没有盲目崇信弦理论人都会把这种状况看得很清楚。理论没能做出任何可以检验的预言，而它的某些支持者不是承认这一点，而是寻求改变规则，使他们的理论不必经历科学通常需要经历的考验。

看来，我们有理由拒绝这种要求，坚持不应仅仅为了挽救一个没能实现我们初衷的理论而改变规则。如果弦理论没有为实验做出独特的预言，如果它除了说我们必须生活在我们能生存的宇宙，而对基本粒子的标准模型没解释任何以前认为神秘的东西，那么它就不大可能成为一个很好的理论。科学史上有过许多起初很有希望而最终失败了的理论。弦理论难道不会是又一个例子吗？

我们很遗憾得出这样的结论：弦论没有做出新的精确的可以证伪的预言。弦论有一些惊人的关于世界的论断。有什么实验或观测能在未来的哪一天找到其中的某个惊人特征的证据吗？就算没有明确的能肯定或否定理论的预言——我们能看到弦的自然观的某个关键特征的证据吗？

弦理论最明显的新奇在于弦本身。如果我们能探测弦的尺度，而

弦理论又是正确的，那么无疑我们能找到大量弦理论的证据。我们会看到很多迹象说明基本对象是一维的而不是点状的。但我们在任何地方都做不了接近所需能量的加速器实验。有什么别的办法能让弦理论自己暴露吗？也许弦会以某种方式变得越来越大，从而我们能看见它们？

最近。科普兰（Edmund Copeland）、迈耶（Robert Myer）和波尔琴斯基提出了这样一个图像。在非常特殊的宇宙学假定下，某些很长的弦也许真的是在早期宇宙生成的，而且今天还存在。[71]宇宙的膨胀现在已经将它们拉伸到了几百万光年那么长。

这种现象并不限于弦论。有段时间，一个流行的星系形成理论也提出它们源于大爆炸留下的电磁流的巨弦。这些所谓的宇宙弦与弦论无关；它们是规范理论结构的结果。它们类似于超导中的量子化的磁流线，可以作为宇宙冷却相变的结果而在宇宙早期形成。我们现在有了来自宇宙观测的确定证据，说明这种弦不是形成宇宙结构的主要成分，但仍然可能存在一些大爆炸留下的宇宙弦。天文学家通过它们对遥远星系的光线的影响来寻找它们。如果宇宙弦来到我们的视线与遥远星系之间，弦的引力场将起着透镜的作用，以特有的方式重复星系的图像。其他事物，如暗物质或别的星系也可以产生类似效应，但天文学家知道如何区分它们产生的像与宇宙弦产生的像。最近有报告说，这种透镜效应可能已经探测到了。它被乐观地标记为CSL-1（宇宙弦透镜1），但从哈勃太空望远镜看，那原来不过是两个靠得很近的星系。[72]

科普兰和他的同事们发现的是，在某些特殊条件下，一根被宇宙膨胀拉得很长的基本弦就像宇宙弦，因此有可能通过透镜效应来观测。这样的基本宇宙弦也可能是引力波的巨大发射体，可以用LIGO（激光干涉仪引力波天文台）看到。

这种预言为我们带来一点希望，也许弦理论将在某一天被观测所

证实。然而宇宙弦的发现本身还不能证明弦理论，因为其他几个理论也预言了这种弦的存在。当然，找不到那根弦也不能否定弦理论，因为那种宇宙弦存在的条件是经过特殊选择的，没有理由认为它们就在我们的宇宙中。

除了弦的存在而外，弦世界还有三个一般性特征：所有合理的弦理论都认同存在额外维，所有的力都统一在一个力，存在超对称。所以，即使我们没有具体的预言，也能发现实验是否能检验这些假定。因为它们独立于弦理论，对其中任何一个的证据都不能证明弦理论是正确的。但反过来就不同了：假如我们知道不存在超对称性、不存在高维或者不存在所有力的统一，那么弦理论就是错的。

我们从额外维说起。也许我们看不见它们，但我们肯定可以寻找它们的效应。方法之一是寻找所有高维理论都预言过的额外力。这些力由构成额外维几何的场传递，这些场肯定是存在的，因为你不可能只让额外维产生我们看到的那些场和力。

我们预计来自那些场的力大致和引力一样强，但在一个或多个方面有别于引力：它们可能具有有限的力程，而且可能不会同等地与所有形式的能量发生相互作用。有些现代实验对这种假想的力很敏感。大约10年前，有个实验发现了这种力的初步证据，被称作第五种力。进一步的实验不支持这个结论，所以那些力到现在也没有证据。

弦理论通常假定额外维很小，但几个大胆的物理学家在90年代意识到，并非一定如此——额外维可以很大，甚至无穷大。这在膜世界的图景中是有可能的。在那样的图景中，我们的三维空间其实是一张膜——就是说，像一张实在的膜，但有三个维——悬浮在有四个或更多空间维的世界里。标准模型的粒子和力——电子、夸克、质子和它们相互作用的力——局限在形成我们世界的三维膜中。所以，仅凭这些力还不能看到额外维的证据。唯一的例外是引力。因为引力无处不在，它能穿越所有的空间维。

这幅图景最先是SLAC（斯坦福线性加速器中心）的三个物理学家阿坎尼-哈姆德（Nima Arkani-Hamed）、达瓦里（Gia Dva-li）和迪莫普罗斯（Savas Dimopoulos）描绘的。他们惊奇地发现额外维可以很大而不与已知实验矛盾。如果存在一个或两个额外维，那么它们可以有毫米级的截面。73

加入这种大额外维的主要效应是，在四维或五维世界的引力将比在三维膜中强大得多，所以，量子引力效应出现的尺度要比人们在其他情形下所预期的长得多。在量子理论中，更长的尺度意味着更小的能量。让额外维达到毫米的长度，就可以降低量子引力效应的能量尺度——从普朗克能量1019 GeV降到1000GeV。这将解决标准模型参数的一个难题：为什么普朗克能量比质子质量大那么多个数量级？但真正令人兴奋的是它把量子引力现象带进了巨型重子对撞机（LHC）即将在2007年揭示的范围。在这些效应中，有可能从基本粒子的碰撞产生量子黑洞。这将是激动人心的发现。

另一幅膜世界的图景是哈佛的兰多尔（Lisa Randall）和约翰霍普金斯大学的桑德鲁姆（Raman Sundrum）描绘的。他们发现，只要更高维世界存在负宇宙学常数，额外维就可以无限大。74值得注意的是，这也符合迄今为止的所有观测，它甚至还预言了新的观测。

这些思想都很大胆，也很有趣，我真佩服它们的创造者。尽管那么说，膜世界的图景却令我困惑。它们也都无奈地面临着以前的高维统一所遭遇过的致命难题。只有当我们特别假定了额外维的几何，特别假定了作为我们世界的三维空间在多维空间的状态，膜世界才可能发生作用。除了旧的卡鲁扎-克莱因理论遭遇的那些问题之外，还有新的问题。如果说有一张漂浮在高维空间的膜，难道不会有很多的膜吗？如果存在其他的膜，那它们碰撞的机会有多少呢？实际上，有人提出大爆炸就是从膜世界的碰撞中产生的。但如果大爆炸能发生一次，为什么后来没发生了呢？已经过去大约140亿年了。也许膜太少了，这样的话，我们又要依赖于精心调节的条件；也许是膜彼此平行

而且没有大的运动，在这种情形，我们仍然是精心调整了条件。

除了这些问题，我的怀疑还在于那些膜图景依赖于特殊的背景几何的选择，而这有悖于爱因斯坦的主要发现，即他在广义相对论中确立的思想：时空几何是动力学的，物理学必须以背景独立的方式来表达。不过，它们终归还算是科学的：思想尽管大胆，但可以用实验来检验。我们还是说得更清楚些吧。如果膜世界的某个预言实现了，它也不足以作为弦理论的证明。膜世界理论是独立存在的；它们不需要弦理论。在弦理论的框架下也不存在完全实现的膜世界的模型。反过来说，如果膜世界没有一个预言实现，这也不能否定弦理论。膜世界只是弦理论的额外维的一种可能表现形式。

弦论的第二个一般性预言是世界是超对称的。这方面仍然没有可以检验的预言，因为我们知道，如果说超对称性真的描述了我们看到的世界，那么它必然是破缺的。我们在第五章说过，超对称可以在LHC中看到。这是可能的，但即使超对称是真的，我们也没有一点儿把握。

幸运的是，还有其他检验超对称的方法。有一种可能的方法涉及暗物质。在许多标准模型的超对称推广中，最轻的新粒子是稳定的，而且不带电荷。这种新的稳定粒子可能就是暗物质。它只能通过引力和弱核力与普通粒子相互作用。我们称这种粒子为WIMP，即大质量弱相互作用粒子。已经有几个实验在探测它们了。这些探测器利用了暗物质通过弱核力与普通粒子发生相互作用的思想。这使那些粒子很像大质量状态的中微子，因为中微子也只通过引力和弱力与物质相互作用。

不幸的是，超对称理论有着太多的自由参数，没能具体预言那些WIMP的质量应该是多少，它们的作用大概有多强。但如果它们确实构成了暗物质，假定它们在星系形成中起着我们设想的那种作用，那么我们就可以推测它可能的质量范围。预言的范围令人满意地落在理

论和实验估计的最轻超伙伴的质量范围内。

实验家们像寻找来自太阳和遥远超新星的中微子那样寻找WIMP，尽管经过了广泛的搜寻，但至今一个也没找到。当然这不是最后的结果——它只不过说明如果它们存在，其相互作用也太微弱，不能引起探测器的反应。我们可以说，假如它们像中微子那样与物质相互作用，我们早就看到了。不管怎么说，只要用任何方法发现了超对称，都将是物理学的辉煌胜利。

我们需要记住的是，即使弦理论要求世界在一定尺度上是超对称的，它也没说那个尺度是多大。因此，如果超对称没有在LHC看到，也不能证明弦理论是错的，因为它所在的尺度是完全可以调节的。另一方面，如果超对称找到了，也不能证明弦理论是对的。有些普通理论也需要超对称，如标准模型的最小超对称扩张。即使在引力的量子理论中，超对称也不是弦理论独有的。例如，量子引力的另一种方法，所谓圈量子引力，也与超对称完全相容。

我们现在来看弦理论的第三个一般性预言：所有的基本力都在某个尺度达到统一。和其他情形一样，这也是一个比弦理论更大的思想，所以它的证明不能证明弦理论是对的；实际上，弦理论允许几种不同形式的统一。但多数物理学家认为，没有一种形式代表了大统一。我们在第三章讨论过，大统一有一个迄今尚未证明的一般性预言，即质子是不稳定的，将以某个时间尺度衰变。很多实验做过质子衰变，都失败了。这些结果（或者说没有结果）排除了某些大统一理论，但没有否定一般的思想。然而，衰变的失败却为可能的理论（包括超对称理论）强加了一种约束。

众多理论家相信这三个一般性预言都将被证明。于是实验家们花了大量精力去寻找可能支持它们的证据。可以毫不夸张地说，在最近30年，成百上千的人和数以亿计的钱都耗费在了寻找大统一、超对称和更高维的迹象。尽管努力了，这些假说的证据一个也没出现。任何

思想的证明，即使不能作为弦理论的直接证据，也将第一次暗示我们，在弦理论要求的那一揽子东西中，至少有某个部分使我们与实在的距离更近了，而不是更远了。

## 第十二章　弦理论解释了什么

我们今天对弦论的异趣都理解了什么呢？第一次超弦革命已经发生20多年了，在这段时期里，弦论吸引了全世界理论物理学的关注和资源——世界上1000多个最有素养、最有才华的科学家在为它工作。虽然有人对理论前景心存疑虑，但科学迟早能获得证据，使我们对某个理论的真实性达成共识。考虑到未来什么事情都可能发生，我想在结束这个部分时，对作为科学理论的一个计划的弦理论做一点评价。

还是说得更明白些吧。首先，我不评价工作的质量；许多弦理论家都是才华横溢而且经验丰富的，他们的工作都是高质量的。其次，我要区分两个问题：弦理论是否是令人信服的物理学理论的候选者？弦理论的研究为数学或其他物理学问题带来了什么有用的认识吗？没人怀疑弦理论引出了很多好数学，我们也深化了某些规范理论的认识。但弦理论带给数学或其他物理学领域的副产品，并不是支持或反驳弦理论作为正确科学理论的证据。

我要评说的是弦论在多大程度上实现了（或将要实现）它最初的允诺：统一量子理论、引力论和基本粒子物理学。弦论或许是爱因斯坦1905年以来的科学革命的顶点。这种评价不可能基于尚未实现的假设或尚未证实的猜想甚至理论追随者们的心愿。这是科学，而一个理论的真实性只能以它在科学文献发表的结果为基础进行评价，因此我们必须谨慎地区分猜想、证据和证明。

也许有人会问，现在做如此评论是否为时尚早。但弦论已经持续发展了35年，而且在20多年里吸引了全世界许多最聪明的数学家。正

如我以前强调的，至少从18世纪末以来，科学史上从来不曾有过一个重要理论，经过了10多年而既没衰落也没获得实验和理论的支持。要说实验难做，也不能令人信服，原因有两点：首先，弦论要解释的多数数据已经存在于宇宙学和基本粒子物理学的标准模型的常数中了。其次，虽然弦确实小到不能直接看见，但以前的理论几乎总能很快引出新的实验——没人想到要做的实验。

另外，在进行评价的过程中，我们需要考虑许多证据。做弦论的人为我们带来了大量需要思考的东西。猜想与假设，虽然经过广泛深入的研究仍然悬而未决，也有着同样的意义。多数未解的关键猜想至少经过10年了，但今天仍然没有能很快解决它们的迹象。

最后，弦论作为第十章讲述的新发现的巨大景观中的一点，正处于危机之中，令许多科学家重新考虑它的前景。于是，尽管我们不会忘记新的发展可能改变这幅图景，现在似乎也该把弦论作为一个科学理论来进行评价了。

任何理论评价的第一步是拿它与实验和观测进行比较。这在第十一章已经讨论了。我们看到，即使弦理论经过了那么多的研究，似乎也不可能用当前可行的实验来证明或否定它的某个独特的预言。

有些科学家会认为这个理由足以抛弃弦理论了，但弦理论是为了解决某些理论难题而创建的。即使没有实验检验，我们也会支持一个能为大问题带来满意解的理论。在第一章里，我描述了理论物理学面临的五个重大问题。能终结爱因斯坦革命的理论应该解决所有这些问题。所以，要公正评价弦理论，应该问它在这方面做得怎么样。

先简要说说我们所知道的弦理论的东西。

首先，它没有完整的形式。弦理论的基本原理是什么？理论的主要方程应该是什么？这些问题还没有公认的意见，甚至没有证明是否

存在这样的一个完整形式。我们对弦理论的认识主要是一些近似结果和与以下四类理论相关的猜想。

1.我们认识最好的理论刻画了在简单背景（如平直十维时空）中运动的弦，其中背景几何不随时间变化，宇宙学常数等于零。也有许多情形，9个空间维中的某些被卷曲而其余空间维平直。这些是我们理解最深的理论，因为可以具体计算在那些背景中运动的弦和膜。

在这些理论中，我们用所谓微扰论的近似方法来描述背景空间中的弦的运动和相互作用。已经证明，这些理论是非常确定的，在近似方法中给出了精确到二阶的有限而和谐的预言。其他结果也支持（不过迄今尚未证明）这些理论的一致性。除此之外，还有大量结果和猜想描述了这些理论中的一个对偶关系网。

然而，这些理论的每一个都与我们世界的事实相矛盾。它们多数都有未破缺的超对称，这在现实世界里还没见过。少数没有未破缺超对称的几个理论预言费米子与玻色子具有相等质量的超伙伴，也不曾见过。除了引力和电磁力外，它们还预言存在无限力程的力，还是没有见过。

2.在具有负宇宙学常数的世界的情形，可以证明存在基于马尔德希纳猜想的一类弦理论。这将在具有负宇宙学常数的特定空间的弦理论与特定的超对称规范理论联系起来了。迄今为止，除了某些非常特殊的、高度对称的极端情形而外，这些弦理论还不能确定地构造出来进行研究。很多证据支持弱形式的马尔德希纳猜想，但还不知道到底哪种形式的猜想是正确的。如果最强形式是正确的，那么弦理论等价于规范理论，这个关系为具有负宇宙学常数的弦理论提供了精确的描述。然而，这些理论也不能描述我们的宇宙，因为我们知道宇宙学常数是正的。

3.人们猜想还存在无数其他理论，相应于更复杂背景下运动的弦，它们的宇宙学常数不等于零，时空背景几何随时间演化，或者其

背景包含着膜和其他场。这包括了大量宇宙学常数为正的情形，与观测结果一致。迄今还不可能精确确定这些弦理论，也不可能进行具体的计算从而导出预言。它们存在的证据是满足一些必要但远非充分的条件。

4.在26个时空维中，有一个理论没有费米子或超对称，叫玻色弦。这个理论有快子，会导致无穷的表达式，造成理论的矛盾。

有人提出，所有这些猜想的和构造的理论都统一于一个更深层的理论，叫M理论。其基本思想是，我们认识的所有理论都将对应于那个深层理论的解。从不同弦理论之间众多的对偶关系（猜想的或证明的）中，可以看到它存在的证据，但迄今还没人能建立它的基本原理，写出它的基本定律。

我们可以从上面的综述看到，为什么弦论的任何评价都必然引起争议。如果我们仅限于考虑已知存在的理论——即可以做具体计算并进行预言的理论——我们必然得出弦理论与自然无关的结论，因为每个弦理论都与实验结果矛盾。所以，我们对弦理论可能描述世界的希望，完全寄托在我们对众多可能存在的弦理论的信念。

不过，许多做弦论的人相信那些猜想的理论是存在的。这种信念似乎基于以下的间接推理：

1.他们猜想存在一般形式的弦理论，由未知的原理和未知的方程决定。这个未知的理论有很多解，每个解都为在某个背景时空里传播的弦提供一个和谐的理论。

2.接着他们写出一组方程，猜想它能逼近那个未知理论的真正方程。然后他们猜想这些近似方程为和谐的弦理论的背景提供了必要但不充分的条件。这些方程可以认为是不同形式的卡鲁扎-克莱因理论，因为它们也包含着广义相对论在高维的推广。

3.他们猜想，对这些近似方程的每个解，都存在一个弦理论，尽管不能具体将它写出来。

这个推理的问题在于，它的第一步是一个猜想。我们并不知道那个理论或决定理论的方程是否真的存在。这就使第二步也成了猜想。同样，我们也不知道猜想的近似方程是否为弦理论的存在提供了充分（与必要相对）的条件。

这类推理存在着风险——它假定了需要证明的东西。如果你相信论证的假定，那么存在性已经隐含在其中的理论就可以作为弦理论来研究。但必须记住，它们不是弦理论，也不是任何形式的理论，而只不过是经典方程的解。其意义完全依赖于没人能建立的理论或没人能证明的猜想。在这种情况下，似乎没有多大的理由相信存在任何尚未确立的弦理论。

从这些能得出什么结论呢？首先，在尚未完全了解弦理论的情况下，有许多可能的特征。根据我们现在的认识，可能会出现某个满足人们愿望的理论。不过，根据我们现在的认识，也有可能不存在真正的理论，所有的东西不过是只有在特殊对称约束下才成立的特殊情形的一系列近似结果。

必然的结论似乎是，弦理论本身——即在背景时空里运动的弦的理论——不会是基本理论。如果说弦理论与物理学有关系，那是因为它为存在一个更基本的理论提供了证据。这是大家公认的，而且那个基本理论有一个名字——M理论——尽管它还没被构造出来。

也许这并不像看起来那么糟糕。例如，从严格意义说，多数量子场论的存在都是不知道的。粒子物理学家研究的量子场论——包括量子电动力学、量子色动力学和标准模型——都和弦理论一样，只有通过近似过程来定义。（尽管在这些情形，至少可以证明它们在所有阶的近似水平上都给出了有限和一致的结果。）而且，有很好的理由相

信，标准模型并不是严格确立的数学理论。但是，只要我们相信标准模型只是通向更深层理论的一步，这一点也没什么好担心的。

起初人们认为弦理论就是那个深层的理论。根据现有的证据，我们必须承认它不是。弦理论和量子场论一样，似乎是一个近似的构造（从它与自然的关系来看），暗示着某个更基本理论的存在。这倒不一定使弦理论不相干，但为了证明它的价值，它必须至少和标准模型做得一样好。它必须预言一些将被证明为正确的新东西，必须解释已经观察到的现象。我们已经看到，它至今还做不了第一件事情。它能做第二件吗？

为回答这个问题，我们来看弦理论对第一章提出的五个关键问题回答得怎么样。

我们先说好的方面。弦理论最初是由第三个问题（即粒子和力的统一问题）激发的，那么作为这样的统一理论它是怎样树立起来的呢？

非常好。在确立和谐的弦理论的背景下，弦的振动囊括了对应于所有已知类型的物质和力的状态。引力子（携带引力的粒子）表现为圈（即闭弦）的振动。引力正比于物体的质量，也就是当然的事实了。光子（携带电磁力的粒子）也来自弦的振动。从我们对强弱核力的认识看，更复杂的规范场也会自然涌现出来。就是说，弦理论一般性地预言了类似的规范场的存在，尽管它没有预言我们在自然界看到的力的特殊混合。

于是——至少在背景时空的玻色子（携带力的粒子）水平上——弦理论统一了引力与其他力。所有四种基本力都从一个基本物体（弦）的振动中产生出来。

那么，玻色子与组成物质的粒子（如夸克、电子和中微子）如何

统一呢？原来，在加入超对称性时，它们也表现为弦的振动状态。这样，超对称弦理论统一了所有不同类型的粒子。

而且，弦理论实现这一切，仅凭一个简单的定律：弦在时空穿越时，它扫过的面积最小。这就不需要任何单独的描述粒子相互作用的定律：弦相互作用的定律直接来自那个描述弦如何运动的简单定律。而描述各种力和粒子的定律也就随之而来了，因为它们都是弦的振动。实际上，根据弦运动满足面积最小的简单条件，我们已经导出了描述力和粒子的传播和相互作用的一整套方程。单独的一种实体，满足单独一个简单的定律——正是这种美妙的简单性，在开始的时候令我们激动，至今还令很多人激动不已。

那么，第一章的第一个问题，量子引力问题呢？情况有点儿复杂。好的方面是携带引力的粒子源自弦的振动，同样，粒子产生的引力正比于粒子的质量。这可以导出引力与量子理论的统一吗？我在第一章和第六章说过，爱因斯坦的广义相对论是背景独立的理论。这意味着空间和时间的整个几何是动力学的；没有一点是固定不变的。引力的量子理论也应该是背景独立的。空间和时间应该由理论产生，而不是作为弦的活动背景。

弦理论目前还没有搭建成一个背景独立的理论。这是它竞选引力的量子理论的主要弱点。我们理解弦理论，是通过弦和其他物体在固定的、不随时间变化的经典空间背景几何中的运动，所以，爱因斯坦关于空间和时间几何是动力学的发现还没有融入弦理论。

有趣的是，除了几个一维理论外，并不存在严格的背景相关的量子场论。所有量子场论都只是用近似过程来定义的。弦理论大概也有这种性质，因为它是背景相关的。我们不禁猜想，任何和谐的量子理论都必须是背景独立的。如果真是那样，就意味着量子理论与广义相对论的统一不是可以选择的，而是必然的。

有人宣称，广义相对论在一定意义上可以从弦理论导出。这是一个重要的论断，它在何种意义上正确，对我们来说是很重要的——一个背景独立的理论如何能从背景相关的理论推导出来呢？一个时空几何是动态的理论如何能从一个要求固定几何的理论推导出来呢？

理由是这样的：考虑一种时空几何，看在那种几何中运动和相互作用的弦是否存在和谐的量子力学描述。当你考察这个命题时，你会发现弦理论和谐的一个必要条件是，在一定近似程度上，时空几何是更高维的广义相对论方程的一个解。所以，广义相对论方程从某种意义说是在弦理论的和谐条件下突现出来的。这是弦理论家声称广义相对论来自弦论的基础。

不过这里有一个陷阱。我刚才讲的是最初的二十六维玻色弦的情形。但我们说过，这个理论具有不稳定的快子，因此并不真的是可行的理论。为了使理论稳定，可以使它成为超对称的。超对称性提出了背景几何必须满足的额外条件。目前，唯一的已知在细节上和谐的超对称弦理论都依赖于不随时间演化的背景时空。[75]因此，在这些情形下，不能说所有广义相对论都归结为超对称性理论的近似。的确，从弦理论得到了很多广义相对论的解，包括所有的既有平直空间也有卷曲空间的解。但这些解都是非常特殊的；广义相对论的一般解描述的是时空几何随时间演化的世界，这才把握了爱因斯坦关于时空几何是动态和演化的思想实质。我们不能只凭那些没有时间相关性的解就说广义相对论是从弦理论推导出来的。我们也不能说拥有了引力理论，因为我们已经看到了许多涉及时间相关性的引力现象。

为回应这些疑问，有些弦理论家猜想，在随时间变化的时空背景下也存在和谐的弦理论，只不过它们太难研究了。据我所知，这些理论不可能是超对称的，也没有具体构造出一个一般的形式。它们存在的证据有两种。第一，有人论证，在为了消除快子而使理论稳定的必要条件中，可以引入少量的时间相关性，而不会破坏那些条件。这种论证似乎有道理，但没有具体构造出来，因而很难判断。第二，人们

揭示了某些特殊情形的细节，但最成功的理论都隐含着时间对称性，因而并不合适。其他理论可能都存在不稳定性问题，或者只有在经典方程水平上的结果，远不足以证明它们是否真的存在。还有些理论具有很强的时间相关性，取决于弦理论本身的尺度。

我们没能在一般的时间相关时空里具体构造出一个弦理论，如果不假定存在某个元理论，我们也不能令人信服地证明弦理论的存在。在这种情况下，我们当然不能断言所有广义相对论都可以从弦理论推导出来。这是另一个开放的问题，需要未来的研究来裁决。

我们还可以问，在弦理论能具体构造出来的那些情形，它是否给出了一个和谐的包容了引力和量子论的理论？就是说，我们是不是至少可以描述微弱如空间几何的涟漪的引力波和力？我们是不是可以用量子理论完全和谐地做到这一点？

在一定的近似程度上，可以做到这一点。迄今为止，尽管已经获得了大量正面的证据，也没有出现过任何反例，但超越近似水平的证明还没有一个完全成功的。当然，弦理论家们普遍相信它是正确的。同时，为了证明它，似乎还面临着巨大的障碍。近似方法（即微扰论）给任何物理问题的答案都是无限多项的总和。对前几项来说，每一项都小于它前面的项，所以，只要将前面几项加起来就能得到近似结果。这是弦理论和量子场论的通常做法。接着，为了证明理论是有限的，还需要证明，对任何可能需要进行的计算，无限多项的每一项都是有限的。

现在的情形是这样的。第一项显然是有限的，但它对应于经典物理学，所以里面没有量子力学的东西。第二项，也就是第一个可能成为无限的项，很容易证明也是有限的。直到2001年人们才完全证明了第三项也是有限的。那是一个壮举，洛杉矶加州大学（UCLA）的德霍克（Eric D‘Hoker）和哥伦比亚大学的蓬（Duong H.Phong）为它付出了多年的心力。[76]然后，他们开始做第四项。他们认识了第四项

的很多东西，但至今也没能证明它是有限的。他们是否能证明所有项都是有限的，还要拭目以待。他们面临的部分问题是，理论的算法在第二项以后就变得模糊不清了，所以他们在证明理论给出有限答案之前，需要先为它找一个正确的定义。

怎么会这样呢？我不是说过弦理论是以一个非常简单的法则为基础的吗？是啊，可问题在于，那个法则只有在用于原来的二十六维理论才算是简单的。当超对称加入进来以后，它就有点儿复杂了。

还有其他结果，它们表明，原本每一项都可能出现某些无限表达，但实际上并没出现。1992年曼德尔斯塔姆发表了一个强有力的证明。最近，一个叫贝科维茨（Nathan　Berkovits）的喜欢在圣保罗工作的美国物理学家有了很大的进展。他构造了一种新形式的超弦理论，得到了有利于微扰论中的每一项的证明，只需要满足几个额外的假定。不过，现在还不好说那些假定是否容易清除。但这仍然是迈向证明的重大进步。有限性问题并没有得到多数弦理论家的关注，而我对少数仍然在这个问题上辛勤工作的人们怀有无限的敬意。

围绕有限性还有一个更令人忧虑的问题。最后，即使每一项的计算都证明是有限的，计算的精确答案需要把所有的项加起来。因为要加无限多项，结果仍然可能是无限的。虽然还没有做过那样的求和，但有证据表明结果将是无限的（问题太专业，不可能在这儿说明白）。换句话说，近似过程只能接近真实的预言，但最后还是偏离了。这是量子理论的普遍特征。它意味着微扰论虽然是有用的工具，却不能用来确立一个理论。

凭现在的证据，没有证明，也没有反例，几乎不可能知道弦理论是否有限。证据可以从两个方面来解读。经过很多艰辛的工作以后（尽管人数很少），有了几个部分的证明。这既可以认为是猜想正确的证据，也可以认为是它存在某些错误的证据。如果说这些天才的物理学家的努力都失败了，如果说每个尝试都不完全，那可能就是因为

他们要证明的猜想是错误的。数学开创证明的思想，将它作为信仰的准则，是因为人类直觉经常会走入误区。普遍相信的猜想有时会证明是错误的。这不是数学严格性的问题。物理学家并不总是像数学家兄弟那样追求严密。大家接受的许多有趣的理论结果并没有数学证明。但我们现在的情形不是那样的。即使就物理学家的严格水平说，弦理论也是没有证明的。

在这种情况下，我不知道超弦理论最终会是有限还是无限。但如果我们认为某个对理论至关重要的东西是正确的，那么就应该花力气将直觉变成证明。的确，我们见过很多流行的猜想，经过了几代人还没有证明，但那通常是因为失去了关键的环节。即使最终证明了人们相信的东西，我们努力的回报往往也是更深入地认识原先滋生猜想的那片数学土壤。

我们以后还会回来讨论为什么弦理论的有限性招惹了那么多争议。现在我们要说的是它并不是一个孤立的例子。激发两次弦论革命的几个关键猜想仍然没有证明，其中包括强弱对偶和马尔德希纳对偶。在两种情形下，都有许多证据表明不同理论之间的某种形式的关系是正确的。即使猜想所称的严格等价性错了，这些思想和结果也是重要的。但从严格方面说，我们必须区分猜想、证据和证明。

有人声称马尔德希纳猜想独立证明了弦理论至少在一定的几何条件下生成了一个优美的引力的量子理论。他们断言，弦理论在某些情形下精确等价于三维空间的某个普通规范理论，生成了在任意阶近似都精确可靠的量子引力理论。

正如我们指出的，这个论断的问题在于，强形式的与尔德希纳猜想尚未证明。有明显的证据表明，在马尔德希纳的十维超对称弦理论与最大超规范理论之间存在着某种关系，但我们目前还没有整个猜想的证明。如果说两个（都没有精确定义的）理论之间只有部分的对应，也很容易解释那个证据。（最近，人们通过所谓格子规范理论的

两次近似方法来逼近那个规范理论，已经取得了令人欣喜的进步。）当前的证据与马尔德希纳的完全等价不存在的猜想是一致的，那是因为两个理论本来就不同，或者因为两个理论严格说来都不存在。另一方面，假如强形式的马尔德希纳猜想是正确的——这同样符合当前的证据——那么弦理论就在具有负宇宙学常数背景的特殊情形下提供了一个良好的量子引力理论。而且，那些理论还是部分背景独立的，就是说三维空间的物理生出了一个九维空间。

还有证据也说明弦理论能提出一个引力与量子理论的统一理论。最强的结果涉及膜和黑洞。这些结果异乎寻常，但正如在第九章说的，它们还走得不够远。眼下它们还仅限于非常特殊的黑洞，要把这些精确结果很快推广到一般的黑洞（包括我们认为自然界存在的所有类型），似乎还很渺茫；而这些结果的出现也许是因为黑洞所具有的额外对称性。最后，弦理论的结果并不包括特殊黑洞的量子几何的具体描述；它们仅限于研究模型的膜系统——这些系统与黑洞具有许多共性，但存在于寻常的平直时空。而且，那些结果是通过近似方法研究的，把引力清除在外了。

有人认为这些极端的膜系统在引力复原时会变成黑洞。但弦理论不可能具体说明如何生成黑洞。真想做到那一点的话，还需要一个在随时间演化的时空背景下的弦理论，而我们知道眼下还没有那样的理论。

自从有了这些关于黑洞的初始结果，后来涌现了大量想象的在弦理论中描述真实黑洞的思想。但它们都遭遇了一个一般性的问题：只要脱离那些可以用超对称进行计算的特殊黑洞，它们就得不到精确的结果。当我们研究普通黑洞时，或者当我们进一步追问奇点发生什么时，我们都不可避免地处于随时间演化的时空几何中。超对称在这儿不灵了，所有依赖它的优美的计算工具也无用武之地了。于是我们也和弦理论的研究一样面临着痛苦的境遇：从特殊情形得到了神奇的结果，却不能确定结果是否可以推向整个理论；也许它只有在我们可以

计算的特殊情形下才是正确的。

面对如此困境，还能说弦理论解决了贝肯斯坦和霍金发现的黑洞熵、温度和信息丢失疑难吗？答案是，尽管有启发性的结果，还是不能说弦理论解决了这些问题。对极端和近似极端的黑洞，运用膜的模型系统进行的计算确实得出了描述相应的黑洞热力学的具体公式。但它们不是黑洞，只是在大量超对称约束下具有黑洞的热力学性质的系统。结果没有为黑洞的量子几何提供真正的描述。所以它们并没有以黑洞的微观描述解释贝肯斯坦和霍金的结果。而且，正如我们看到的，结果仅适用于非常特殊的黑洞类型，而不能用于具有实际物理意义的黑洞。

概括说来：基于眼下的结果，我们不能自信地声称弦理论解决了量子引力问题。证据很杂乱。在一定近似下，弦理论似乎和谐地统一了量子理论和引力，给出了合理而有限的结果。但难以确定这是否适于整个理论。有证据支持马尔德希纳猜想，但猜想本身尚未证明，而只有当整个猜想正确了，我们才敢说有了一个良好的量子引力理论。黑洞的图景很迷人，但那仅限于弦理论能模拟的非典型黑洞。除此之外，还有一个老生常谈的问题：弦理论不是背景独立的，即使在局限意义下，它也只能描述几何不随时间演化的静态背景。

我们只能说，在这些局限下，有那么一些证据表明弦理论预示着某个和谐的引力与量子理论的统一。但弦理论本身就是那个和谐的统一吗？可能不是，因为它没有解决那些问题。

现在来看第一章列举的其他问题。第四个问题是要解释粒子物理学标准模型的参数值。弦理论目前显然还没能做到这一点，也没有理由相信它能做到。相反，正如我们在第十章讨论的，证据表明和谐的弦理论实在太多了，几乎不可能在这个问题上提出什么预言。

第五个问题是解释暗物质和暗能量、解释宇宙学常数。情况也不

容乐观。弦理论通常包含着比观测更多的粒子和力，因而确实提供了很多暗物质和暗能量的候选者。某些额外的粒子可以是暗物质，某些额外的力可以是暗能量。但弦理论没有明确预言在众多可能的候选者中，到底哪个是暗物质或暗能量。

例如，在可能的暗物质候选者中，有一种叫轴子的粒子（其名称指特定的性质，就不多说了）。[77]许多（但不是所有）弦理论都包含着轴子，起初看来这是好事。但多数包含轴子的弦理论都预言轴子具有与标准宇宙学模型相矛盾的性质，于是那又成了坏事。可还有很多弦理论，虽然同样包含轴子，却能和宇宙学模型一致。而且，在这一点上，宇宙学模型也可能是错的。因此有理由认为，如果轴子是暗物质，那么这和弦理论是一致的。但这决不等于说弦理论预言了暗物质是轴子，也不等于说它做出了额外的预言，从而就能凭暗物质的观测来证伪它自己。

剩下的问题是第二个：量子力学的基本问题。弦理论为这些问题提出了什么答案吗？没有。关于量子理论的基础问题，弦理论目前还没有任何直接的说法。

当前的情形就是这样。弦理论潜在地彻底解决了五大问题中的一个，即粒子和力的统一问题。这是弦理论产生的动力，迄今仍然是它最动人的成果。

有证据说弦理论为解决量子引力问题指明了方向，但它顶多预示着存在一个更深层的能解决量子引力问题的理论，而不能说它就是那个理论。

当前，弦理论还没有解决其余的三个问题。它似乎不可能解释物理学和宇宙学的标准模型的参数。它为暗物质和暗能量提出了一系列可能的候选者，但不能唯一预言或解释它们的任何东西。直到今天，弦理论也无言面对那伟大的奥秘，即量子理论的意义。

除了这个，就没有什么成功可言了吗？我们寻求一个理论的成功，往往是看它对新实验或新观测做出了什么预言。我们说过，弦理论绝对没有这类预言。它的力量在于统一了我们知道的所有类型的粒子和力。举例说，假如我们不知道引力，我们也可以从弦理论预言它的存在。这是很重要的，但不是对新实验的预言。而且，即使我们找到了某个不满足预言的实验或观测，也不能因此而证明弦理论是错的（即证伪）。

如果说弦理论没有新预言，那么我们至少也应该问问，它对我们知道的数据是如何解释的。情况很特别。由于我们的知识不够完备，我们不得不将许多可能的弦理论划分为两组，并分别考察每一个组。第一组由已知存在的弦理论组成，第二组包括那些猜想的但尚未构造出来的弦理论。

最近发现了宇宙膨胀在加速，我们被迫关注第二组的弦理论，因为只有它们符合新的观测。但我们不知道如何计算弦在这些理论中运动和相互作用的几率。我们也不能证明这些理论确实存在。我们关于它们的证据是它们的背景满足某些必要但绝非充分的条件。所以，即使在最有利的情况下，即使存在描述我们宇宙的弦理论，也需要创造新的技术来计算那些新理论中的实验预言。正如我们看到的，已知的弦理论都不满足我们看到的世界：多数理论有未破缺的超对称；其余的则预言了等质量的费米子与玻色子成对出现；而所有理论都预言存在新的具有无限作用范围的力（迄今尚未观测到）。由此难免得到一个结论：不论动机多么美好，弦理论还是没能实现人们在20多年前所寄予的希望。

在1985年的鼎盛时期，有个新革命理论的最热烈支持者，叫弗里丹（Daniel Friedan），当时在芝加哥大学费米研究所。可是在最近一篇论文里，他也不得不说：

作为一个物理学理论，弦理论失败了，因为存在多个可能的背景

时空……弦理论长久以来的危机是它完全不能解释或预言任何大距离的物理。弦理论不能决定宏观时空的维度、几何、粒子谱和耦合常数。弦理论不能明确地解释现实世界的现有知识，不能做出任何具体的预言。弦理论的可靠性无法估计，更不可能确立。作为候选的物理学理论，弦理论毫无可信度。[78]

不过，还是有许多弦理论家知难而上。可是，面对我们讨论的那些问题，怎么还有那么多聪明人在不懈地为弦理论工作呢？

令弦理论家们倾心的一点是理论的美妙或“雅致”。这是一种美学判断，也许人们不会赞同，我也不知道该如何评价它。不管怎么说，它与理论成就的客观评价无关。正如我们在第一部分看到的，很多美妙的理论最后都证明是与自然无关的。

有些年轻的弦理论家提出，即使弦理论不能作为最终的成功的统一理论，它还是产生了一些能帮助我们理解其他理论的副产品。他们特别谈到了第九章讨论过的马尔德希纳猜想，它就提供了一种方法，使我们能通过在对应的引力理论中的简单计算来研究某些规范理论。这对具有超对称的理论来说当然行之有效。但是，假如要它在标准模型中发挥作用，就必须用于没有超对称的规范理论。这时还可以用其他技术，问题是马尔德希纳猜想比它们好多少。结论还没有。有一个好的检验例子，是一个简化的只含两个空间维的规范理论。它最近已经有了精确解，不过所用的技术与超对称或弦理论毫不相干。[79]还可以用另一种方法来研究它——通过计算机的繁重的计算。计算机的计算是可靠的，因此可以作为比较其他方法的预言的标准。这样的比较表明，马尔德希纳猜想并不比其他技术好。[80]

有的理论家还将数学的潜在进步作为继续做弦理论的理由。其中一个可能的进步涉及六维空间几何，那是弦理论家们当作可能的紧化空间维的例子来研究的。在某些情形下，弦理论的数学预言了这些几何的意外的惊人的性质。这是令人欣喜的，但我们应该清楚到底发生

了什么。它们与物理学没有关系。它们是纯粹发生在数学平面的事情：弦理论提出了联系不同数学结构的猜想。弦理论认为，六维几何的性质可以表达为更简单的数学结构，它们可以定义于弦在时间中扫过的二维曲面上。这种结构叫共形场。他们猜想，某些六维空间的性质是那些共形场理论的结构的镜像。这引出六维空间之间的惊人关系，是弦理论的一个绝妙的副产品。但它的应用并不需要我们将弦理论看作关于自然的理论。首先，共形场在许多不同的应用（包括凝聚态物理学和圈量子引力）中都起着重要作用，所以它和弦理论没有什么独特的关系。

弦理论还在另一些情形引出了数学发现。在一种非常美妙的情况下，弦理论的某个玩具模型（叫拓扑弦理论）使我们对高维空间拓扑学有了新的认识。然而，这个证据本身并不能证明弦理论就是正确的自然理论，因为拓扑弦理论是弦理论的一种简化形式，没有统一自然存在的粒子和力。

更一般地说，即使一个物理理论激发了数学的进步，也不能以此证明它就是正确的物理理论。错误的理论也曾激发过许多数学进步。托勒密的本轮理论就刺激了三角和数论的发展，但它并不因此而正确。牛顿物理学激发了数学的主要部分的发展，但如果牛顿理论与实验相矛盾，则什么也救不了它。很多理论都建立在美妙的数学基础上，但从来没有成功，也从来没有人相信。开普勒关于行星轨道的第一个理论就是一个令人警醒的例子。所以，尽管一个研究纲领可能激发出美妙的数学猜想，但那也挽救不了一个没有清晰的核心原理和物理学预言的理论。

弦理论面临的困境植根于整个统一理论的基础。在本书第一部分，我们认识了令早期统一理论困惑的巨大障碍——它们最终导致了统一的失败。其中有些是想通过引入更高的空间维来统一世界。后来发现那些更高维的几何远非唯一的，而且随不稳定性崩溃了。我们在前面的章节看到，其基本原因就在于统一总会产生结果，这意味着存

在新的现象。在有利的情形下——如麦克斯韦的电磁理论、温伯格和萨拉姆的弱电理论、爱因斯坦的狭义和广义相对论——新现象很快就发现了。这是我们可以为统一而欢呼的难得的几个例子。在其他的统一尝试中，新现象没有很快发现，或者与观测矛盾。这些统一的结果当然没人欢呼，而理论家们还必须设法将那些结果隐藏起来。我没见过这种隐藏结果的统一最后能引出好理论来；这些统一迟早是会被抛弃的。

不论超对称还是高维空间，到头来都不得不费很多气力来隐藏它们的统一的结果。结果发现，超对称没有联系任何两个已知的粒子；相反，每个已知的粒子却有一个未知的超对称伙伴。而为了要那个未知的粒子成为不可见的，必须调整理论的很多自由参数。在高维情形，理论的所有解几乎都与观测相矛盾。少有的几个揭示了某种事物（如我们的世界）的解却是众多可能解的海洋中的不稳定小岛，看起来全然是不相容的东西。81

弦理论能摆脱这些降临在早期高维理论和超对称理论中的问题吗？看样子不能，因为它需要隐藏的东西远远超过了卡鲁扎-克莱因理论和超对称理论。斯坦福小组提出的稳定高维的建议也许有效，但代价太高，产生了一片猜想的解的景观。于是，为了避免卡鲁扎-克莱因理论的致命问题，其代价是至少需要采纳弦理论家们从一开始就拒绝的观点，即我们必须同等地将众多可能的弦理论都看作自然的可能描述。这意味着人们原来盼望的唯一的因而能为基本粒子物理学提出可证伪预言的统一，不过是一个幻影。

在第十一章讨论过，苏斯金、温伯格和其他一些人声称弦理论景观也许是物理学家的前景，我们发现这些论断是不能令人信服的。那么它会将我们引向何处呢？在最近的访谈中，苏斯金指出，我们的风险在于，要么接受弦景观和它所隐含的科学方法的淡化，要么放弃整个科学，拿神的智能设计来解释标准模型里的参数选择：

假如由于某种意外的原因——如数学的原因，或者因为它与观测的矛盾——[弦理论的]景观成了不和谐的东西，我敢肯定物理学家将继续寻求世界的自然解释。但我不得不说，假如真是那样，我们就会处于十分尴尬的境地。如果不解释自然为什么会那样精密地调节参数，我们将被迫面对智能设计的批评。也许有人会说，我们对出现数学的唯一解的期待和智能设计一样，都是基于一种信仰。82

但这是错误的选择。我们很快就会看到，还有其他理论为五大问题提出了真正的解答，它们正在迅速地进步着。废弃弦理论并不意味着废弃科学，而是意味着废弃一个曾经令人欢喜却没能达成人们愿望的方向，为的是将精力集中在其他更有希望成功的方向。

弦理论在很多方面都成功了，自然令人希望它的某些部分或其类似的东西能构成未来的理论。但也有确凿的证据表明它出问题了。自20世纪30年代以来人们就明白了量子引力理论必须是背景独立的，但为了构造这样一个能描述自然的弦理论，依然没有什么进展。同时，对自然的唯一的统一理论的追求，却引出了可能存在无限多个理论的猜想，其中没有一个能具体地写出来。相应的，无限多个理论引出无限多个可能的宇宙。除此之外，我们所能详尽研究的所有形式的理论都不符合观测事实。尽管有大量诱人的猜想，也没有证据说明弦理论能解决理论物理学的那几个大问题。相信猜想的人发现，他们所处的智慧宇宙大不同于那些坚持只相信证据的人们的宇宙。在正统的科学领地里存在分歧如此巨大的观点，这个事实本身就说明存在重大的问题。

那么，弦理论还值得研究吗？或者，是否应该（如某些人所想的）宣布它的失败呢？许多希望破灭了，许多关键的猜想尚未证明，这两点大概足以使有些人放弃弦理论的研究。但它们还不是完全终止研究的理由。

假如未来某个时候有人找到了某个方法，能构造一个能唯一引向

粒子物理标准模型的弦理论，而且是背景独立的，就存在于我们看到的三维非超对称世界中，那会怎么样呢？即使发现这种理论的前景很渺茫，也是有可能的——多样的研究纲领对科学的健全来说总是好事，我们以后还要谈这一点。

所以弦理论家的方向当然也是值得走下去的。但还能继续认为它是理论物理学家的主导范式吗？旨在解决理论物理学关键问题的多数资源还该继续支持弦理论的研究吗？其他研究方法还该为了弦理论而继续荒芜吗？只有弦理论家才有资格享有令人羡慕的工作和研究经费吗（这正是目前的情形）？我想这些问题的答案肯定都是否定的。在任何水平上，弦理论都没有成功到那样的程度，值得将所有的蛋都放进它的篮子里。

如果没有其他方法值得我们去做呢？有些弦理论家鼓吹支持弦理论是因为它是“唯一的选择”。我要说，即使真是这样，我们也应该热情鼓励物理学家和数学家去探索不同的方法。假如没有别的思想，那我们就创造一些出来。因为弦理论在近期内还无望产生可以证伪的预言，所以也没有什么特别急切的事情。还是让我们鼓励人们寻求一条捷径来回答理论物理学的那五大问题吧。

事实上，真有其他的方法——旨在解决那五大问题的其他理论和研究纲领。尽管多数理论家关心着弦理论，也有少数人在其他领域取得了巨大的进步。更重要的是，他们还有新实验发现的线索，是弦理论所未曾预料的。一旦得到证实，它们就将为物理学指明新的方向。这些新理论和实验的发展是本书下一部分的主题。

# 第三篇　弦论之外

## 第十三章　真实世界的惊奇

希腊哲学家赫拉克利特给我们留下一句美妙的格言：自然喜欢隐藏。这真是千真万确的。赫拉克利特没有办法看见原子。不论他的追随者们对原子如何玄想，要看到一个原子，已经远远超出了他们所能想象的技术水平。如今，理论家们大大发挥了自然不可预测的倾向。如果说自然真是超对称的或具有更高的空间维，那么她已经将它很好地隐藏起来了。

但有时候恰好相反。关键的东西就摆在我们面前，等着大家去看。躲过赫拉克利特的视线的东西，在我们今天看来是很容易觉察的，已经习以为常了，如惯性原理或自由落体的不变加速度。伽利略关于地球运动的观测也用不着望远镜或机械钟。在我看来，它们早在赫拉克利特时代就应该发现了。他只需要提出正确的问题。

于是，当我们哀叹难以检验弦理论背后的思想时，我们应该问问哪些东西隐藏起来了。在科学史上，有许多发现令科学家惊讶，因为它们出乎理论的预料。今天是不是也有理论家不曾寻求过、理论也不曾预言过的东西呢？它们也许能将物理学引向一个有趣的方向。会不会我们已经看到了它们，却因为它们的存在有碍我们的理论过程而被忽略了呢？

答案是肯定的。最近有几个实验结果预示着多数弦理论家和粒子物理学家都未曾想到的新现象。这些现象都还没完全确定。有几个情形的结果很可靠，但解释有分歧；其他情形的结果则因为过于新奇而

没得到大家的认可。[83]不过还是值得在这儿描述一下，因为假如其中任何一个线索成了真正的发现，那么基础物理学将显现任何形式的弦理论都没预言并难以与之相容的重要特征。这样，其他方法将别无选择地成为基本方法。

我们从宇宙学常数说起，一般认为它代表了加速宇宙膨胀的暗能量。第十章说过，暗能量是弦理论和多数其他理论所不曾预料的，我们也不知道如何确定它的数值。很多人为它苦苦思索了多年，但还是一片茫然。我也不知道答案，不过我有一个设想。我们暂且不考虑用已知的知识来解释宇宙学常数的值。假如我们不能凭已知的东西来解释某个现象，这大概就预示着我们需要寻找新的东西。也许宇宙学常数就是某个新东西的征兆，在那种情形它大概还有别的表现。我们该如何去寻找它们、认识它们呢？

答案很简单，因为普遍现象终归是简单的。物理学中的力只要几个数字来刻画——例如，力的传播距离和决定其强度的力荷。刻画宇宙学常数的是尺度，即它令宇宙卷曲的距离尺度。我们称这个尺度为R，大约等于10亿光年（即1027 cm）。[84]宇宙学常数的怪异在于它的尺度远大于物理学的其他尺度。R是原子核大小的1040倍，普朗克尺度（大约是质子大小的10-20）的1060倍。所以人们自然想知道尺度R是否代表了某种全新的物理。为此，寻求发生在同样巨大尺度的现象，应该是一个好办法。

宇宙学常数的尺度上发生了什么吗？我们从宇宙学本身说起。我们最精确的宇宙学观测是对宇宙微波背景辐射的测量。这是大爆炸留下的辐射，它从遥远太空的各个方向到达我们。它纯粹是热辐射——就是说，它是随机的。随着宇宙的膨胀，它已经冷却下来了，现在大约是2.7K。这个温度在整个天空都是非常均匀的，只有大约十万分之几的涨落（图13-1上）。涨落的状态为极早期宇宙提供了重要信息。

过去几十年里，微波背景的温度涨落已经通过卫星、气球探测器

和地面探测器勾画出来了。为了理解这些实验的测量结果，可以将涨落看作宇宙早期的声波。接着再看不同波长的涨落有多大。结果是一幅图像，如图13-1下，它告诉我们不同的波长所具有的能量。

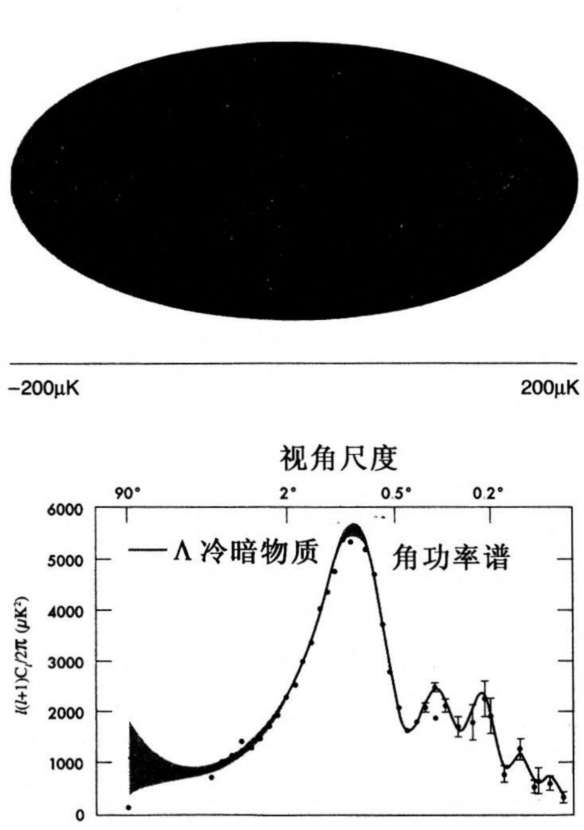



图13-1上：从微波频率看到的天空。来自我们银河系内部的信号已经被清除了，留下的宇宙图像是它刚好冷却到电子和质子结合成氢的时候。下：上图在不同波长的能量分布。点代表WMAP和其他来源的数据，曲线是标准宇宙模型预言的拟合

图131有一个主峰，跟着几个小峰。这些峰值的发现是当代科学的一大胜利。根据宇宙学家的解读，它们说明早期宇宙的物质处于共振状态，就像鼓槌或长笛。乐器振动的波长正比于乐器的大小，宇宙也是如此。共振态的波长向我们揭示了宇宙第一次透明时的大小：那是大爆炸后30万年左右，原初的等离子“退化”或“解耦”成为分离的物

质和能量，微波背景变得可见了。这些观测对确定宇宙学模型的参数是极端重要的。

我们从数据看到的另一个特征是最大波长的能量很小。这也许只是统计涨落，因为它包含的数据量比较小。但假如那不是统计的原因，就可以解释为一种截断，超过它就很少有激发的模式了。有趣的是，这个截断就在与宇宙学常数相关的尺度R。

从广为接受的极早期宇宙理论（即暴胀）的观点看，存在这样的截断是令人疑惑的。根据暴胀理论，宇宙在极端早期指数式地膨胀。暴胀解释了宇宙背景辐射的近似均匀性。它的解释是在于确证我们现在看到的宇宙各部分在宇宙还充满着等离子的时代可能就已经是因果关联的了。

理论还预言了宇宙微波背景的涨落，而它们曾被假定是暴胀时期的量子效应残余。不确定性原理意味着在暴胀宇宙的能量中起主导作用的场应该是涨落的，这些涨落都印在了空间的几何中。当宇宙指数式膨胀时，它们持续涨落着，在宇宙透明时产生的辐射温度也跟着涨落。

暴胀可能生成一个具有相对均匀性质的巨大宇宙区域。根据尺度的简单论证，这个区域比可见区域要大许多个数量级。如果暴胀恰好在那个区域和我们现在看到的区域一样大的时刻停下来，那么在暴胀物理学中一定存在某个参数，才可能选择一个特殊的停止时刻，那正好就是我们的时代。但这几乎是不可能的，因为暴胀发生的时候，宇宙的温度比今天最热的恒星中心的温度还高10到20个数量级。因此，暴胀的定律一定是不同的，只能主导那种极端的条件下的物理。关于暴胀的定律有很多假设，但没有一个涉及100亿年的时间尺度。换句话说，当前的宇宙学常数值似乎不可能与引起暴胀的物理学有任何关系。

因此，如果说暴胀产生了我们看到的均匀宇宙，那么它很可能产生了一个在更大的尺度上均匀的宇宙。这意味着暴胀产生的涨落模式，不论我们看得多远，都应该一直延续下去。如果你能看到观测宇宙以外的地方，你仍然应该看到宇宙微波背景的涨落。但数据表明涨落在尺度R就可能停止了。

实际上，宇宙学家在考察微波背景的大尺度涨落模式时，还发现了更多的疑问。宇宙学家们普遍相信，宇宙在最大尺度上是对称的——就是说，任何方向都是彼此相同的。看到的情形却并非如此。辐射在那些大尺度上的模式不是对称的，而存在一个特殊的方向。（宇宙学家兰德（Kate Land）和马古维约（Joao Magueiyo）称它是“魔轴”。）[85]还没有人为这个现象提出过合理的解释。

这些观测之所以引发争议，是因为它们完全违背了我们基于暴胀的期待。因为暴胀解释了很多宇宙学问题，很多谨慎的科学家怀疑微波数据可能有错。实际上，测量完全有可能是错误的。数据发表之前经过了大量仔细的分析，其中之一就是剔除已知来自我们星系的辐射。这一步可能做得不对，但熟悉数据分析过程的科学家几乎都不相信会出现那种情况。还有一种可能是，我们的观测只不过是统计异常。尺度R的某个波长的振动占据了大部分空间——大约60度；于是我们只看到了很少的波长，因而只有很少的数据，所以我们看到的可能只是随机的统计涨落。如果说存在某个特殊的方向是统计反常，估计它出现的几率小于1%。[86]但人们大概更容易相信那个不太可能的运气，而不愿相信暴胀预言的失败。

这些问题眼下还没有解决。不过对现在来说，我们知道这一点就够了：我们在R尺度寻找奇异的物理现象，果然找到了。

还有与R尺度相关的其他现象吗？我们可以结合R和其他自然常数，看看在引出的新尺度上会发生什么。举一个例子，考虑R除以光速：R/c。这是一个时间量，大约是宇宙今天的年龄。它的倒数c/R是

一个频率——“音调”非常低，相当于宇宙的一生才振动一次。

下一个最简单的量是c2/R，是一个加速度。它其实是宇宙膨胀的加速度——就是说，是由宇宙学常数引起的加速度。然而，它和寻常的加速度相比却小得可怜：$10^{-8}$（cm/s2）。看一只在地板上爬行的小虫子，它大约每秒爬10cm。假如它在一只狗的一生时间里将速度加倍，那么它的加速度就是C2/R，当然是很小的。

不过我们可以假定存在一种新的能解释宇宙学常数值的普遍现象。根据尺度相当的事实，新现象应该也能影响任何其他具有如此小加速度的运动。于是，每当我们看到任何事物以这样的小加速度运动，就可以期待看到新的现象。事情于是变得趣味盎然了。我们确实知道一些加速如此缓慢的事物。一个例子就是绕着典型星系旋转的典型恒星。一个星系环绕另一个星系的加速甚至更慢。那么，这样的小加速度的恒星轨道与更大加速度的恒星轨道是不是有什么不同的地方呢？答案是肯定的，我们确实看到了，而且差别很大。这就是暗物质问题。

我们在第一章讨论过，天文学家是通过测量恒星相对于星系中心的轨道加速度发现暗物质问题的。之所以产生这个问题，是因为天文学家可以根据观测的加速度推测星系物质的分布。在大多数星系中，结果与直接观测到的物质相矛盾。

现在我可以更详细地说说偏差出现在什么地方（为简单起见，我只讨论螺旋星系，其中多数恒星在盘状的圆形轨道上运动）。在发现问题的每个星系中，只有在一定轨道以外运动的恒星才受影响。而在那个轨道之内则没有问题——那里的加速度和可见物质引起的一样。所以，星系内部似乎存在一个区域，其中牛顿定律依然成立，而不需要暗物质来帮忙。在那个区域以外，事情就麻烦了。

关键的问题是：分离两个区域的轨道在什么地方？我们可以假定

它出现在距离星系中心的某个特殊位置，这是自然的假定，却是错误的。那么分界线是不是处于一定的恒星或光线密度呢？这个答案也是错的。奇怪的是，决定分界线的似乎正是加速度本身。随着距离星系中心越来越远，加速度将越来越小，存在某个临界的加速度，它标志着牛顿引力定律的崩溃。似乎只要恒星加速度超过那个临界值，牛顿定律就成立，预言的加速度就等于我们看到的加速度。这种情形不需要任何暗物质。然而当观测的加速度小于那个临界值时，它就不再满足牛顿定律的预言了。

那个特殊的加速度等于多少呢？测量结果大约是$1.2\times10^{-8}$（cm/s2）。这个值恰好接近宇宙学常数预言的加速度c2/R！

暗物质故事里的这个不寻常转机是一个叫米尔格罗姆（Mordehai Milgrom）的以色列物理学家在20世纪80年代初发现的。他在1983年发表那个发现，但多年来一直被忽略了。[87]然而，随着数据的改进，他的发现越发显得正确。尺度C2/R刻画了牛顿定律在星系的什么地方失败。天文学家们现在称它为米尔格罗姆定律。

我要让你们明白这个发现有多奇怪。尺度R是整个观测宇宙的尺度，比任何一个星系的尺度都大得多。我们已经看到，加速度c2/R就出现在这个宇宙学尺度，它是宇宙膨胀的加速度。这个尺度完全没有理由影响单个星系的动力学。但在观测数据面前，我们不得不承认它确实有影响。我还记得第一次听说它时是多么惊讶。我惊呆了，也激动了。我茫然地转悠了一个钟头，嘟囔着脏字眼儿。实验终于说话了！世界的秘密比我们理沦家们想象的多得多！

这要怎么解释呢？除了巧合而外，还有两种可能。一是可能存在暗物质，尺度C2/R可以刻画暗物质粒子的物理学；或者尺度c2/R可以描述星系的暗物质晕，因为它与暗物质坍缩形成星系时的密度有关。不论哪种情形，暗能量与暗物质都是不同的现象，不过二者是有联系的。

另一种可能是，不存在暗物质而牛顿引力定律在加速度小于特殊值C2/R时失败。在这种情形，需要新的定律来代替这种条件下的牛顿定律。米尔格罗姆在1983年的论文里提出了这样一个理论。他称之为MOND，即“修正的牛顿动力学”。根据牛顿引力定律，物体由于某个质量产生的加速度会随着远离那个质量而以特殊的方式减小——即随距离的平方减小。米尔格罗姆的理论指出，牛顿定律只有在加速度减小到那个魔幻的数值$1.2\times10^{-8}$ cm/s2之前才能成立。小于那个数值时，引力不随距离的平方减小，而只随距离反比例地减小。另外，通常的牛顿力正比于引起加速度的质量乘以一个常数（牛顿引力常数），而MOND说的是，当加速度很小时，力正比于质量的平方根乘以牛顿常数。

如果米尔格罗姆是对的，那么特殊轨道以外的恒星之所以加速更快，是因为它们经历着比牛顿预言更强的引力！这是崭新的物理学——不在普朗克尺度下，甚至不在加速器里，而就在我们面前，在我们看到的天空的恒星的运动中。

作为一个理论，MOND对物理学家没多大意义。引力和电力随距离的平方而减小，有着很好的理由。那是相对论与空间的三维特征相结合的结果。我不能在这儿说得太详细，但结论是强有力的。米尔格罗姆的理论似乎背离了基本的物理学原理，包括狭义和广义相对论的原理。

很多人尝试过修正相对论来构造一个包容MOND或类似东西的理论。贝肯斯坦构造了这样一个理论；莫法特（John Moffat，当时在多伦多大学）也构造了一个，另外还有康涅狄格大学的曼海姆（Philip Mannheim）。他们都是极富想象力的人（你大概还记得第六章说过贝肯斯坦，他发现了黑洞熵；莫法特也发明了很多惊人的东西，包括可变光速宇宙学）。三个理论都在一定程度上有用，但在我看来，它们太人工化了，一点儿也不自然。它们具有几个额外的场；为了满足观测，还要求将几个常数调节到不太可能的数值。我还担心

理论的稳定性问题，尽管作者们声称问题已经解决了。好消息是，人们可以用老方法来研究这些理论——将它们的预言与我们掌握的大量天文学观测数据进行对比。

应该说，MOND在星系外的表现并不太好。我们有很多大于星系尺度的星系质量分布和运动的数据。在这种情况下，暗物质理论比MOND对数据的解释要好得多。

尽管如此，MOND似乎在星系内部表现很好。[88]过去10年获得的数据表明，在已经研究过的80多种情形（据最近的统计），MOND精确预言了恒星是如何运动的。实际上，MOND比基于暗物质的理论更好地预言了恒星在星系内部的运动。当然，暗物质的理论也一直在进步，所以我还不敢预言它们较量的结果会如何。但是现在我们似乎面临着一种喜忧参半的状况。我们有两个迥然不同的理论，其中只有一个可能是对的。一个理论——基于暗物质的理论——感觉很好，很容易令人相信，很好预言了星系外的运动，但对星系内的情形则不是太好。另一个理论，即MOND，在星系内的表现很好，在星系外失败了，而其假定则无论如何似乎总是与已经确立的科学针锋相对。我承认，在最近一年里，没有任何问题像这个问题一样令我寝食难安。

如果不是因为米尔格罗姆定律提出神秘的宇宙学常数尺度与恒星在星系的运动多少有些关系，人们很容易忽视MOND。仅从数据看，加速度c2/R似乎对恒星运动起着重要作用。不管这是因为暗物质与暗能量之间的深层联系，还是因为某种更基本的东西，我们都看到可以在这个加速度发现新的物理学。

我和我认识的几个最有想象力的理论家讨论过MOND。情况通常是这样的：我们总是谈某个严肃的主流问题，而老有人说起星系。我们会相视一笑，于是有人说，“看来你也担心MOND了，。”仿佛在对暗号。接着，我们共享疯狂的思想——因为所有关于MOND的思想，如果当时看不出错误来，都是疯狂的。

唯一的好处是，这种情形有很多数据，而数据越来越好。我们迟早会知道是真的存在暗物质，还是应该接受对物理学定律的彻底修正。

当然，暗物质和暗能量有相同的物理学尺度，也许只是巧合。并非所有巧合都有意义。所以，我们要问是否还有其他能测量那个微小加速度的现象。如果有，会不会出现理论与实验矛盾的情况？

看来，确实存在那种情形，而且同样令人不安。宇航局（NASA）迄今已向太阳系外发射了几艘飞船。其中的两艘——先驱者10和先驱者11——行几十年了。这些“先驱者”是为外行星旅行设计的，它们在太阳系的平面上沿着和行星相反的方向运行，离太阳越来越远。

NASA在加州帕萨迪纳喷气实验室（JPL）的科学家们可以根据多普勒频移确定先驱者飞船的速度，从而发现它们的精确轨道。JPL还想通过预言太阳、行星和太阳系的其他事物作用在两艘飞船的力来预告它们的轨迹。在两种情况下，观测的轨迹都不符合预言的结果。[89] 误差来自额外的将飞船拉向太阳的加速度。那个神秘加速度的大小大约是8×10-8（m/s2）——大约是在星系中测量的反常加速度的6倍。不过，考虑到两个现象之间没有明显的关联，两个数值还是相当接近的。

我要说的是，人们对这种情形的数据还没有完全认可。虽然两个先驱者都发现了反常，这比一个反常更令人信服，但它们都是JPL制造和跟踪的。然而，JPL数据是由科学家们用太空合作的高精度卫星运动程序独立分析的，结果都和JPL一致。所以，数据至今还是可靠的。但天文学家和物理学家有更高的证明标准（这是可以理解的），更何况我们现在正面临着牛顿引力定律可能在太阳系外失败的问题。

由于加速度偏差很小，也许是因为某个小小的效应（例如，飞船

向阳的一面会比背阴的一面稍微热一点儿），或者因为气体泄露。JPL小组考虑了他们能想到的每种可能的效应，至今也不能解释观测到的反常加速度。最近，有人提出发射一个特殊设计的探测器，尽可能清除那些乱真的效应。这样的探测器还要等多年才能飞出太阳系，但即使如此，这件事也是值得一做的。牛顿引力定律已经确立300多年了，哪怕需要更多的年月来证明或否定它，也是不足为怪的。

如果MOND或先驱者反常是正确的，又将怎样呢？它们的数据能与某个现有的理论相容吗？

MOND与迄今研究过的所有形式的弦理论无论如何是不相容的。那么它能与某个未知的弦理论相容吗？当然。由于弦理论的多变，这种可能是无法排除的，尽管也很难实现。其他理论又如何呢？有几个人费了很大力气，想从膜世界图景或某种形式的量子引力导出MOND。思想是有了几个，但都不令人满意。我在圆周理论物理研究所的同事马科普洛（Fotini Markopoulou）和我曾考虑从量子引力得到MOND，但不能具体说明我们思想的功用。MOND是一个诱人的神秘理论，但现在还不能求解。所以，我们还是来看看从其他实验生出的新物理的线索。

最动人的是那些彻底颠覆人们普遍信仰的实验。有些信仰深深嵌入我们的思想，也表现在我们的语言。例如，我们说物理学常数，是指那些永不变化的数。它们包括物理学定律的最基本参数，如光速或电子电荷。但这些常数真的不变吗？为什么光速不能随时间变化呢？能探测那样的变化吗？

在第十一章讨论的多宇宙理论中，我们假定参数在不同宇宙间变化，但我们怎么才能在自己的宇宙中观测那样的变化呢？那些常数（如光速）会在我们的宇宙中随时间变化吗？有的物理学家指出，光速是在某个单位制下测量的——即每秒多少千米，那么，在单位系统本身随时间变化的情况下，你怎样识别光速的变化呢？

为回答这个问题，我们必须知道距离单位和时间单位是如何定义的。这些单位建立在一定的物理学标准上，而那些标准是通过某些物理学系统的行为来定义的。首先是参照地球的标准：1m原来等于从地球北极到赤道距离的百万分之一。现在的标准以原子性质为基础——例如，1秒是用铯原子的振动频率来定义的。[90]

如果考虑单位的定义，物理学常数就定义为比值。例如，只要我们知道了光穿过原子的时间和原子发射光的周期之比，就可以定义光速。这些比值在所有单位制都是相同的。这些比值纯粹取决于原子的物理性质，它们的测量不涉及单位的选择。由于比值纯粹是用物理性质确定的，所以有理由追问它们是否随时间变化。如果是，那么原子的一种性质与另一种性质之间的关系也将随时间而改变。

比值的变化可以通过原子发射的光的频率的变化来测量。原子发射出离散频率的光，构成光谱，因此这些频率将产生许多比值。我们可以问，这些比值对那些来自遥远恒星和星系的光——即数十亿年的光——是否有什么不同。

这类实验没能探测到自然常数在我们星系或临近星系内的变化。就是说，在百万年的时间尺度上，常数没有任何可以感知的变化。不过，澳大利亚的一个小组正在进行的实验，从来自类星体的光——百亿年前的光——发现了那些比值的改变。澳大利亚科学家们没有去研究类星体本身的原子光谱；他们做的更聪明。光线从类星体来到我们，一路上穿过了众多的星系。每当它穿过一个星系，就有部分光被那个星系的原子吸收。原子吸收特定频率的光，但由于多普勒效应，被吸收的光的频率向光谱的红端移动了一定的量，那个量正比于星系到我们的距离。结果，来自类星体的光谱由大量谱线组成，每一根谱线对应着一定距离外的星系所吸收的光。通过研究这些光线的频率之比，我们可以发现基本常数在光从类星体到达我们的时间里所发生的变化。因为变化必然表现为频率之比，而基本常数有几个，于是物理学家决定研究最简单的比值——由决定原子性质的常数组成的精细结

构常数。这个常数叫α，等于电子电荷的平方除以光速、乘以普朗克常数。

澳大利亚人利用夏威夷Keck望远镜拍摄的精确光谱，研究了来自80个类星体的光线。他们从那些数据得到，大约100亿年前，α比今天小万分之一。[91]

这变化很小，但如果确实，那么它将是一个重大发现，是几十年来最重要的发现。这是人们第一次发现基本的自然常数在随时间发生改变。

我认识的许多天文学家都怀着开放的态度。总的说来，数据经过了非常认真仔细的分析。没人发现澳大利亚小组的方法或数据有什么明显的缺陷。但实验本身太精密了，精度恰好在可能的边缘，我们不能排除分析里存在某种误差。我写本书时，情况还很混乱。其实任何新实验技术都是这样的。别的小组在做着同样的测量，结果却有争议。[92]

许多理论家都怀疑这些精细结构常数的变化证据。他们担心这种变化太不自然，因为它将在电子、核子和原子理论中引入一个时间尺度，比原子物理学尺度小很多个数量级。当然，他们可以对宇宙学常数的尺度说同样的话。实际上，除了宇宙学常数本身之外，精细结构常数发生改变的尺度并不接近任何已经测量过的东西。所以，这可能是与尺度R有关的另一个神秘现象。

尺度R还有一个表现，大概就是神秘的中微子质量。我们可以用物理学的基本常数将长度尺度转化为质量尺度，结果它与不同种类的中微子质量差有相同数量级。没人知道为什么中微子（最轻的粒子）会具有和R有关的质量，但它确实有了——这是另一条诱人的线索。

最后还有一个有关尺度的实验。结合实验与牛顿引力常数，我们

可以得出一个结论：可能存在某些效应能改变毫米尺度的引力。目前，华盛顿大学阿德尔博格（Eric Adelberger）领导的一个小组正在非常精确地测量分离几毫米的两个物体间的引力。到2006年6月，他们也只能公开宣布，在6%mm的尺度下他们尚未发现牛顿定律错误的证据。

即使没有别的结果，应该说我们的实验至少是检验了物理学的基本原理。人们普遍认为这些原理一旦发现，就是永恒不变的，但历史却不是这样的。几乎每个所谓的基本原理都被取代了。多数原理，不论多么有用，多么接近自然现象，随着探索自然世界的实验越来越精确，迟早都会失败的。柏拉图声称天球上的万物都沿圆周运动，这是有很好理由的：月亮天球上的万物被认为是永恒而完美的，而最完美的莫过于圆周上的匀速运动。托勒密采纳了这个原理，用它进一步构造了本轮——沿圆周运动的圆周。

行星轨道其实非常接近圆，行星在轨道上的运动也几乎是匀速的。有趣的是，行星轨道中最不圆的是水星的轨道——而它也非常接近圆，视力最好的人才勉强能分辨出它与圆的偏差。1609年，开普勒经过9年的艰苦工作，终于认识到水星的轨道是椭圆。那年，伽利略将望远镜对着天空，开创了天文学的新纪元，最终发现开普勒是正确的。圆是最完美的图形，但行星轨道不是圆。

古人宣称圆是最完美的形状时，意思是它是最对称的：轨道的每一点都与其他点相同。这样的原理最难以割舍，它们满足了我们对对称的需求，并将观测到的对称性提升为必要的条件。现代物理学以一组对称性为基础，它们无疑装点了最基本的原理。许多现代理论家和古人一样，本能地相信基本理论必须是最对称的可能定律。我们应该相信直觉，还是该汲取历史的教训呢？历史告诉我们（正如行星轨道的情形一样），我们看得越近，自然就变得越不对称。

最深植根于当代理论的对称性来自爱因斯坦的狭义和广义相对

论。其中最基本的是惯性坐标系的相对性。那其实就是伽利略原理，自17世纪以来成为物理学的基本思想。它只是说我们不能区分静止与速度和方向都不变的运动。正是因为这个原理，我们才感觉不到地球的运动或我们在匀速飞行的飞机上的运动。只要没有加速度，你就不能感觉自己的运动。换句话说，这意味着没有优越的观察者，也没有优越的坐标系：只要没有加速度，所有的观察者都是一样的。

爱因斯坦在1905年做的，就是将这个原理用于光。结果是，我们必须认为光速为常数，与光源或观察者的运动无关。不论你我如何相对运动，我们都会赋予光子以相同的速度。这是爱因斯坦狭义相对论的基础。

在狭义相对论下，我们能做出许多关于基本粒子物理学的预言。有个预言是关于宇宙线的。有一组穿过宇宙的粒子（主要是光子），它们到达地球大气的顶部，在那儿与空气中的原子发生碰撞，变成粒子簇射，像雨一样落下来，可以在地面探测。谁也不知这些宇宙线的来源，但它们能量越高，越是罕见。我们曾看到它们有比质子质量高1000亿倍的能量。为了具有如此高的能量，质子必须以非常接近光的速度运动，大约是0.9999999999个光速——根据狭义相对论，没有粒子能超过光速。

人们相信宇宙线来自遥远的星系；如果真是那样，它们在到达我们之前大约经过了数百万甚至数十亿光年。1966年，两个苏联物理学家扎泽宾（Georgiy Zatsepin）和库兹闵（Vadim Kuzmin），以及康奈尔大学物理学家格莱森（Kenneth Greisen），仅用狭义相对论，分别独立做出了一个惊人的关于宇宙线的预言。[93]他们的预言，即通常所谓的GZK预言，值得多说两句，因为它现在还在经受考验。这是有史以来对狭义相对论的最极端考验。实际上，它是第一次对狭义相对论在接近普朗克尺度的考验，我们有可能在那个尺度看到量子引力理论的效应。

优秀的科学家善于发挥所有工具的作用。格莱森、扎泽宾和库兹闵就意识到，我们已经接近了一个巨大的实验室，它比地球上所能建造的任何实验室都大得多，那就是宇宙本身。我们可以探测在经历了大部分宇宙年龄的旅行之后到达地球的宇宙线。在它们的旅行中，非常微弱的效应——在地球实验中不可能看到的小效应——可以放大到我们能看到它。如果用宇宙作为实验工具，我们可以看到比人们从前想象的要深刻得多的自然结构。

关键的一点在于，宇宙线穿越的空间不是真空，而是充满着宇宙微波背景辐射。格莱森和苏联科学家们意识到，能量大于某个特定值的质子将与背景辐射中的光子发生相互作用，产生新的粒子（很可能是π子即π介子）。这种粒子生成过程需要能量，而能量是守恒的，于是高能光子会慢下来。这样看来，如果质子携带的能量超过生成π子需要的能量，那么空间对它来说就像是不透明的。

于是，空间的作用仿佛是一种过滤器。构成宇宙线的质子，只有在能量低于生成π子需要的能量时，才能穿过。如果能量太高，它们会生成π子，减慢速度，不断重复相同的过程，直到不再能生成π子。这就像宇宙为质子规定了一个速度极限。GZK三人预言，达到地球的质子的能量，都不会超过以这种方式生成π子所需要的能量。他们预言的生成π子所需要的能量大约是普朗克能量（1019 GeV）的十亿分之一，被称作GZK截断。

这是很大的能量，比我们所知的任何事物都更接近普朗克能量。它比当前规划的最精密的粒子加速器所能产生的最高能量还要高1000万倍。GZK预言是对爱因斯坦狭义相对论的严峻考验。与地球上做过的或可能做的任何实验相比，它在更高的能量、更接近光速的速度的情况下检验了相对论。1966年，当GZK预言发表时，人们只看到过能量低于预言的截断能量的宇宙线，不过当前已经制造了几个仪器，可以探测等于甚至超过截断能量的宇宙线粒子。其中一个实验叫AGASA（明野巨型空气簇射阵列），是在日本进行的，报告了至少十

多起那样的极端事件。这些事件的能量大于$3\times1020$ eV——大约相当于棒球手击球的能量，却是一个质子携带的。

这些事件也许预示着狭义相对论在极端能量下失败了。哈佛物理学家科尔曼（Sidney Coleman）和格拉肖在20世纪90年代末提出，狭义相对论的破产可能提高生成π子所需要的能量，从而提高GZK截断能量，预示更高能量的质子也能达到我们地球的探测器。[94]

这并不是对观测到的高能宇宙线质子的唯一解释。它们也可能是在距离地球足够近的地方生成的，还没来得及与宇宙微波背景发生相互作用，因而速度还没有减慢。这只需要看看那些质子是不是来自天空的某个特殊地方。目前还没有证据，但是很有可能。

这些极端高能的粒子也可能完全不是质子。它们可以是迄今未知的某些稳定粒子。本来就比质子的质量大得多。如果是这样，那也是一个重大发现。

当然，这些实验也可能是错的。AGASA小组报告说，他们的能量测量中有25%的不确定性，这是一个很大的误差来源，但还不足以解释他们看到的高能事件的存在。然而，他们对实验精度的估计也许是错误的。

幸运的是，正在进行的一个实验将解决这个矛盾。这就是奥格（Auger）宇宙线探测器，正运行在阿根廷西部的彭巴斯草原。如果奥格探测器证明了日本人的观测，如果可以忽略其他可能的解释，那么它将是最近百年来最重大的发现——人们将眼睁睁地看到，构成20世纪科学革命的一个基础理论第一次失败了。

为了观测如此极端能量的宇宙线粒子，需要什么工具呢？具有那么高能量的粒子来到大气顶部时，会产生其他类型的粒子簇射，像雨一样降落到一片广阔的土地。奥格实验由分布在阿根廷彭巴斯草原

3000km2的几百个探测器构成。同时，几个高精度光学探测器不断扫描天空，捕捉粒子簇射产生的光。通过综合所有这些探测器的信号，奥格的研究者们就能决定原来那些落在大气层的粒子的能量，判断它们来自什么方向。

我写这些东西时，奥格天文台刚公布第一批数据。好消息是实验运行良好，但仍然没有足够的数据判定基于狭义相对论预言的截断是否存在。不过我们还是有理由希望经过几年的运行之后，会有足够数据来解决这个疑问。

即使奥格小组宣布狭义相对论依然成立，其发现本身也将是最近25年来——即未发现质子衰变（见第四章）以来——最重要的基础物理学发现。理论在没有实验指导下的黑暗中摸索的漫长岁月终于到头了。但如果奥格发现狭义相对论不完全正确，那么它将预示着新物理学的到来。那时，我们还需要花一定的时间来弄清那个革命发现有什么意义，会把我们引向什么地方。

## 第十四章　站在爱因斯坦肩头

假如说奥格计划或其他什么实验证明爱因斯坦的狭义相对论破产了，这对弦理论来说是一个坏消息：它意味着21世纪的第一个重大实验发现竟然全然出乎最流行的“万物之理”的预料。弦理论假定狭义相对论是正确的，和爱因斯坦在100年前写下它的时候一样。实际上，弦理论的主要贡献之一就是构造一个与量子论和相对论都协调一致的关于弦的理论。所以，弦理论预言，不论不同频率的光子来自多么遥远的地方，它们都以相同的速度传播。我们已经看到，弦理论没做出多少预言，但这是一个；其实，它是弦理论唯一的一个可以用目前技术检验的预言。

如果狭义相对论的预言错了，又将意味着什么呢？有两种可能。

一种可能是狭义相对论错了，而另一种可能是进一步深化它。因为这一区别，引出了最近几十年来基础物理学中最惊人的新思想。

有几个实验可能粉碎或修正狭义相对论。奥格实验能做到，但我们对γ射线爆发的观测也能做到。那是一种剧烈的爆发，它在若干分之一秒内产生的光，和整个星系发出的光一样多。顾名思义，多数这样的光都是γ射线的辐射，它们是能量较高的一些光子。大概平均每天都有一个这样的爆发信号来到地球。第一个信号是在20世纪60年代末由军事卫星（本来是为了寻找非法的核试验）发现的。现在有专门的科学卫星在观测它们。

尽管有一些可能的理论，但我们还不知道γ射线爆发是什么来源。它们可能来自两颗中子星的碰撞或一颗中子星与一个黑洞的碰撞。不论哪种情形，两个天体都应该相互环绕几十亿年了，但这样的系统是不稳定的。当它们以引力波形式辐射能量时，它们会非常缓慢地盘旋着相互靠近，直到最终发生我们所知的最剧烈的碰撞。

爱因斯坦的狭义相对论告诉我们，光不论频率多少，都以相同速度传播。γ射线爆发为检验这个论断提供了实验条件，因为它们在很短的时间内爆发出多种能量的光子。最重要的是，它们可以经过数十亿年才到达我们，因此走进了实验的核心。

假如爱因斯坦错了，不同能量的光子将以略微不同的速度传播。如果两个在相同距离处产生的光子在不同时刻到达地球，那么这无疑预示着狭义相对论失败了。

如此重大的发现有什么意义呢？这首先依赖于理论失败所在的物理学尺度。我们预料狭义相对论可能失败的一个尺度是普朗克长度。回想一下第十三章说的，普朗克尺度大约是质子大小的10-20。量子理论告诉我们，这个尺度代表了一个临界点，在小于它的尺度下，经典的时空图景将彻底瓦解。爱因斯坦狭义相对论是那个经典图景的一

部分，所以我们可以认为它将在那个尺度崩溃。

有什么实验能看到空间和时间结构在普朗克尺度破裂的效应吗？在当前的电子学水平，可以探测不同光子到达我们的微小时间差，但电子学是不是足以测量更微弱的量子引力效应呢？几十年来，理论家们一直在说，普朗克尺度太小，目前能做的实验还不能探测到它。就像100年前多数物理学教授认为原子太小而看不到，我们也在无数的论文和讲义里重复着这个谎言。那真是一个谎言。

值得注意的是，直到20世纪90年代中期，我们才认识到其实我们是可以探测普朗克尺度的。和许多事情一样，当少数几个人认识到它，想发表他们的思想时，却招惹了一片嘘声。其中一个是西班牙物理学家冈萨雷斯一梅斯特（Luis Gonzalez-Mestres），在巴黎国家科学研究中心工作。像这样的发现，在某人拿去公开发表之前，都会被不同的人独立发现过多次。在这个例子中，另一个发现者是罗马大学的乔万尼·阿梅林诺-卡梅里亚（Giovanni Amelino-Comelia）。他现在40出头，有着意大利南方人所特有的魅力和热情，钟情于物理，全身心地投入物理。量子引力的科学家们很幸运有这样一个伙伴。

当乔万尼在牛津做博士后时，就下决心寻求一种方法来观测普朗克尺度。这在当时似乎完全是疯狂的野心，但他敢于证明常识是错误的，并找到了证明的方法。他从质子衰变的检验获得了灵感。预言的质子衰变（见第四章）是一种极端稀有事件，但假如把足够多的质子放在一起，就可能看到它的发生。巨大的质子数量起着放大的作用，使极端微小和稀有的事件成为可见的。乔万尼给自己提出的问题是，这样的放大作用能否使他探测普朗克尺度的现象。

我们已经看到了两个放大的例子：宇宙线和来自γ射线爆发的光子。两种情形下，我们都把宇宙本身作为一个放大器。它的巨大尺度放大了极端稀有事件的几率，而光子经过的漫长时间可以放大微弱的效应。人们早就指出这些实验可能在理论上预示狭义相对论的失败。

乔万尼发觋，我们的确能设计出探测普朗克尺度（也包括量子引力）的实验。

量子引力引起的典型的光速变化简直小得令人难以置信，但来自γ射线爆发的光子可以经过数十亿年的旅行，从而将这个效应放大了。几年前，根据量子引力效应的粗略估计，物理学家们计算不同能量的光子在经历那么长的旅行后，到达我们的时间间隔大约是1/1000秒。这是短暂的时间，但完全落在现代电子学的测量范围内。实际上，最新的γ射线探测器GLAST（γ射线大域太空望远镜）已经具备了这种灵敏度。它计划在2007年夏天发射，人们热切期待着它的结果。

乔万尼和他的合作者们第一次打破壁垒以来，我们已经发现了许多用具体实验探测普朗克尺度的方法。乔万尼的疯狂问题已经成为人们认可的科学领域。

让我们设想一下，假如某个新实验结果在普朗克尺度与狭义相对论冲突，那么它会告诉我们什么有关空间和时间本性的东西呢？

我在本章开头就说过，有两种可能。我们已经讨论了一种，即运动的相对性可能是错误的——意味着我们可以区分绝对运动与绝对静止。这将颠覆自伽利略以来的已成为物理学关键的一个原理。我个人认为这种可能是令人厌恶的，但作为科学家，我必须承认那确实是可能的。其实，如果日本的宇宙线实验AGASA的结果成立，那就已经说明我们看到了狭义相对论的这种失败。

但这是唯一的可能吗？多数物理学家可能会说，如果不同能量的光子以不同速度传播，那么狭义相对论就是错的。十年前我当然也会说这样的话。但我可能错了。

爱因斯坦的狭义相对论基于两个假定：第一个是运动的相对性，第二个是光速的不变性和普适性。会不会第一个假定对了而第二个假

定错了？如果那是不可能的，爱因斯坦就不会硬提出两个假定。可是我认为，直到最近人们才意识到，只改变第二个假定也能得到一个和谐的理论。结果真的可以，但认识这一点却是我在职业生涯中有幸亲身经历的最激动人心的事情之一。

新理论叫修正的或双狭义相对论，简称DSR。它来自一个似乎会引出悖论的简单问题。

我们已经说过，普朗克长度被认为是一个界限，小于它的尺度将出现一种新的、本质上是量子力学的几何。不同的量子引力方法都有一点共识：普朗克长度在某种意义上是可以观测的最小尺度。问题是，所有观测者都同意什么是最小长度吗？

根据爱因斯坦的狭义相对论，不同观测者所看到的运动物体的长度是不同的。和米尺一起的观测者会说尺子是1m长，但任何相对于它运动的观测者看到它要短一点儿。爱因斯坦称之为长度收缩现象。

但这意味着不可能存在所谓“最小长度”之类的东西。不论多短的长度，你总能通过接近光速的相对运动使它变得更短。这样，普朗克长度的概念与狭义相对论之间就存在着矛盾。

现在，你可能认为卷入这个量子引力问题的专家们都被这个矛盾挡住了。你甚至可能认为聪明的大学生刚读一年级物理时就能提出这个问题。毕竟，在弦理论和量子引力中做着最艰难工作的杰出的物理学家，都是从天真的学生走过来的。难道就没有几个看出这个问题吗？就我所知，几乎没有，直到最近。

看出那个问题的是乔万尼。1999年，他遇到了刚才说的那个疑惑，然后解决了它。它的思想是拓展爱因斯坦走向狭义相对论的路线。

狭义相对论的第二个假定（光速是普适的）似乎更是自相矛盾。

为什么呢？考虑两个观测者跟踪一个光子。假定两个观测者相对运动。如果他们测量光子的速度，我们通常认为会得出不同的结果，因为那是正常的物体行为。例如，我们看一辆从身边超过的公共汽车，在我看来它的速度是10km/h；因为我的小汽车以140km/h的速度奔跑，所以站在路旁的观测者会看到公共汽车的速度是150km/h。但是，假如我在同样状况下观测一个光子，狭义相对论告诉我们，路旁的观测者将看到那个光子的速度和我看到的一样。

那么，为什么这不是一个矛盾呢？关键在于，我们没有直接测量速度。速度是一个比值，它是一定时间经过的一定距离。爱因斯坦的核心认识是，不同观测者，即使以不同速度相对运动，测得的光子总是具有相同的速度，因为他们测量的空间和时间都不同。他们测量的时间和距离的变化方式，恰好满足光速是一个普适的量。

可是，为什么这对光是不变的，对其他东西却不是呢？我们不能对距离也玩儿同样的技巧吗？就是说，一般说来，我们知道，观测者测量运动的米尺没有1m长。这对多数长度都是正确的，但是，当我们一路下来，直到普朗克长度时，那效应就会消失了吗？这意味着如果尺子恰好是普朗克长度，那么即使它在运动，所有观测者也会得到相同的长度。那么，我们是不是有了两个普适的量呢？一个速度，一个长度。

爱因斯坦对速度的技巧成功了，是因为没有什么东西能比光跑得更快。世界上有两类事物——以光速运动的事物和以低于光速运动的事物。如果一个观测者看到某个事物比光慢，那么所有观测者都一样；如果一个观测者看到某个事物和光一样快，那么所有观测者也都一样。

乔万尼的思想是对长度运用同样的逻辑。他提出修正空间和时间测量随不同观测者变化的法则，使它满足，如果观测对象是普朗克长度，则所有观测者都认同它具有普朗克长度；如果比它更长，则所有

观测者也有同样的结果。这个纲领可以是和谐的，因为对任何观测者来说，没有比普朗克长度更短的东西了。

乔万尼很快发现，爱因斯坦狭义相对论方程的一种修正可以实现这种想法。他称之为双狭义相对论，因为创立狭义相对论的技巧在这儿运用了两次。我曾沿着他的思路寻找探测普朗克尺度的方法，但他在2000年向大家散发他双狭义相对论思想的论文时，我起先还弄不明白呢。95

那是很恼火的事情，但还有更令人恼火的。大约10年前，我也陷入了同样的困惑。困惑来自我正在研究的圈量子引力，那是引力的一种量子理论。细节并不重要——关键在于我们的圈量子引力计算似乎和爱因斯坦的狭义相对论相冲突。现在我明白了，这些特殊计算实际上真的和爱因斯坦狭义相对论矛盾。但那时候，这种情形想起来就令人恐慌，经过思想斗争以后，我放弃了整个研究路线。实际上，这是最终令我放弃圈量子引力而做弦理论的系列步骤的第一步。

但就在我放弃它的时候，我有了一个想法：也许可以修正狭义相对论，使它满足所有观测者（不论运动与否）都有同样的普朗克长度。这是双狭义相对论的关键思想，尽管我没有足够的想象力为它做任何事情。我想了一下，看不出有什么意义，就去做其他事情了。虽然10年后看到了乔万尼的论文，却没能让我想起过去。我只好从其他方向来把握这个思想。当时我是伦敦帝国学院的访问教授，在那儿认识了一个著名物理学家，叫若昂·马盖若（Joao Magueijo），是来自葡萄牙的年轻宇宙学家，和乔万尼的年纪差不多，也洋溢着同样的拉丁式的热情。

马盖若的名声在于他有一个真正疯狂的思想：光在极早期宇宙中传播更快。这个想法使暴胀成为多余，因为它解释了早期宇宙的每个区域是如何能有因果关联从而达到相同温度的。不需要极早时期的指数式膨胀，也能产生这样的结果。

结果不错，但想法太疯狂了——真正的疯狂。它与狭义和广义相对论都格格不入。除了说它“异端”，恐怕找不到别的字眼儿了。然而，英国的科学界对异端很宽容，马盖若在帝国学院成长起来了。如果他在美国，我想有着那种思想的他未必能做博士后。

马盖若和帝国学院年轻的阿尔布里希（Andreas Albrecht）教授一起发展了他的思想。还在宾夕法尼亚大学读研究生时，阿尔布里希就是暴胀理论的创立者之一。最近，他离开英国回到了美国。我在帝国学院待了几个月后，才发现他和我是同路人。他想知道是否有方法使他的可变光速（VSL）宇宙学思想与狭义和广义相对论一致。不知为什么他觉得和我交谈可能会有所帮助。

我那时并不知道事情已经有人做了。实际上，整个VSL宇宙学更早就由多伦多大学那位想象力丰富的物理学教授莫法特发展起来了。经历了多次“异端”，莫法特发现了他的思想，并以与狭义和广义相对论协调一致的方式解决了它，但他想在专业杂志上发表论文时，却被拒绝了。

马盖若在2003年的《比光速还快》一书中告诉我们，当他和阿尔布里希正打算发表自己的论文时，听说了莫法特的工作。[96]这时，他表现了一贯的作风，热情地和莫法特交了朋友——现在他们的关系还很密切。他开始与我交谈时，已经知道了莫法特的工作，但我想他还没有理解人家已经解决了他正想解决的问题；或者他已经知道了，但是不喜欢那种解决方法。

莫法特如今是我在圆周理论物理研究所的朋友和同事，他的胆略和创造力实在令我钦佩不已。我也曾说过我是多么欣赏乔万尼对探测普朗克尺度的见识，可我得痛苦地承认，若昂和我忽略了他们两位的工作。从某种意义说，我们做对了，因为我们发现了不同的方法，一样可以协调可变光速与相对性原理。如果我知道问题已经解决了——不是一次，而是两次，肯定不会再为它费力气。

若昂经常带着这个问题来找我。我也总是找时间和他交谈，因为他的活力和独特的物理学眼光已经吸引了我。但在几个月里，我都没认真考虑过他所说的。当他拿一本老书给我，看见里面也讨论过那个问题，我才转变了思想。那是著名俄罗斯数学物理学家福克（Vladimir Fock）写的一本广义相对论教科书。[97]我（和所有物理学家一样）了解一些福克在最子场论的工作，但我从没见过他关于相对论的书。若昂想要我考虑的问题是福克书中的一道家庭作业题。当我看到问题时，顿时就想起我十几年前的想法，于是整个事情就豁然开朗了。其实，问题的关键就是保留爱因斯坦的狭义相对论原理，但要改变它的观测法则，使所有观测者所看到的光速和普朗克尺度都是普适的。实际上，常数的速度不再是所有光子的速度，而只是能量很低的光子的速度。

起初我没看出问题和这个思想有什么关系。我们有了一点数学，但还没有完整的理论。大约就在那时，我做了一次旅行，要在罗马停留。我在那儿和乔万尼谈了几个钟头。我突然明白了他在说什么。他早就有了我们正在探究的思想，而且第一个解决了它。不过，他的解决方法里还有很多我不明白的东西。数学看起来很复杂，好像关系着波兰数学物理学家小组在十几年前建立的某种形式——那是我肯定不可能精通的。

我花了很多年去理解那个问题的数学细节。直到我读了英国数学家马吉德（Shahn Majid，量子群的创立者之一）的早期论文，才透彻地理解了。他的工作与波兰数学家小组用的数学有着密切的联系。马吉德从几个想象的概念开始，讲我们应该如何在单一的数学结构下表达相对论和量子论的基本发现。由此他发展了量子群（这是对称性思想的革命性扩展），然后在我们所谓的非对易代数的基础上修正了相对论。他的发现是清楚表述DSR所需要的数学核心，但至少对我来说，在第一次读他复杂的论文时，并没看出这一点。

无论如何，若昂和我都忽略了数学，而只是一味地谈物理。2001

年9月，我移居加拿大，加入新建的圆周理论物理研究所，我们的进展被迫停下来了。一个月后，若昂来到研究所，成为它的第二个访问学者。他到达的那天下午，理论终于尘埃落定了。我们在滑铁卢小区的一个叫“会饮”的咖啡馆里，坐在舒适的躺椅上工作。他还没倒过时差，而我刚从9·11事件后的纽约度周末回来，疲惫不堪。若昂讲话时，我都睡着了，醒来时发现他也在打瞌睡。我还记得他在我失去意识时说的一些话，在便笺纸上写写画画，然后又睡着了。他开始讲话时，我又醒了，又有了几分钟大家都清醒的时间。那个下午就在我们的交谈、计算和瞌睡中过去了。我不知道咖啡馆的服务员会怎么看我们。但我们在某个时刻突然想到了一个几个月都没想到的关键因子，它与位置和动量的交换有关。当我们精疲力竭时，已经发现了DSR的第二种形式，比乔万尼发展的形式简单得多。它就是专业人士现在所知的DSRII。

这大概就是若昂向往的东西。在我们的形式中，能量越高的光子跑得越快。于是，在极早期的宇宙，当温度很高时，光速也很高，总的说来都高于今天的光速。当我们回到更远的过去，温度接近普朗克能量，光速成为无穷大。需要更长的时间，人们才发现这引出一种与广义相对论原理也和谐一致的可变光速理论，但我们终于还是发现它了。我们借平钦（Thomas Pynchon）小说的名字，称这个理论为引力虹。98

“双狭义相对论”是个很笨拙的名字，但固定下来了。它的思想很优美，已经有很多人研究和讨论过。我们不知道它是否描述了自然，但我们对它有足够的认识，知道它是有可能的。

人们最初对DSR的反应不是欢欣鼓舞。有人说它自相矛盾，有人说它只不过是爱因斯坦狭义相对论的一种复杂写法，还有人则两方面都批判。

为了回答第二个批评，我们证明了那个理论做出了不同于狭义相

对论的预言。参与这些讨论的关键角色是一个来自华沙的重金属音乐发烧友，名叫耶日（Jerzy Kowalski-Glikman）。（也许只有欧洲人能真正充当这两种角色。）我相信他第一个真正理解了乔万尼在说什么。我肯定是先明白了他的短小精悍的论文，才读懂了乔万尼的论文——连篇累牍，印得密密麻麻，尽是旁白和细节。耶日发现了双狭义相对论的几个重要结果，也是他清理了我们的工作与他的波兰同事们先前的数学工作之间的关系。

一天下午，我们几个人在多伦多我女朋友的家里展开了一场讨论，我才豁然明白了DSR，知道了它的不同方法是如何相互联系的。乔万尼、耶日、若昂和我紧紧地围坐在狭窄餐厅的小桌旁，试图深入我们分歧和误会的根底。耶日平静地主张，一样东西要有意义，就必须满足一个和谐的数学结构，对他来说，那就是他和他的波兰同事们研究过的非对易几何。若昂说，与物理学有关的任何事物都可以离开虚幻的数学来认识。乔万尼指出，如果不关心哪些数学表达对应于可以测量的东西，我们对这些理论当然可以随便乱说。有个时候——我忘了那是什么话引起的——乔万尼抓起锋利的面包刀吼道，“如果你说的是对的，我就割断我的喉咙。就现在！”

我们盯着他，吓得不敢说话，片刻过后，我们突然放声大笑，他也笑了。从那以后，我们才开始倾听彼此在说什么。

实际上，DSR有不同的形式，也有不同的预言。在某些预言中，存在一个不可逾越的能量，就像最大光速一样。在其他预言中，没有最大能量，但有最大动量。这是很不幸的，因为这削弱了理论的预言能力；但它似乎也无损于理论的和谐，所以我们还得把它留下来。

为了说明DSR的和谐，我们可以证明它在某个可能的宇宙中是正确的。那个可能的宇宙很像我们的宇宙，区别在于它的空间只有二维。20世纪80年代，人们发现量子引力可以精确定义在只有两个空间维的世界。我们称它为2＋1量子引力，代表两个空间维和一个时间

维。而且，如果没有物质，理论还可以精确求解——就是说，我们可以找到精确的数学表达来回答有关理论所描述的世界的任何问题。

结果发现，DSR在任何具有两个空间维、量子引力和物质的世界里都是正确的。人们确立的DSR的特殊形式是乔万尼原来发现的形式。当耶日和我回顾文献时，发现有几个人已经看到了这个二维世界与DSR有关的一些特征，但那时还没有DSR的概念。我们很兴奋，向圆周研究所的同事弗雷德尔（Laurent Freidel，从法国来的，做量子引力研究）讲了这个情况。他告诉我们，他不但已经知道，而且早就想告诉我们了。我相信那是真的。在讨论中，弗雷德尔比我更有精神，我常常听不懂他讲的东西，于是他讲得更快、更大声。不管怎样，我们还是合写了一篇文章，解释为什么DSR对二维空间的宇宙一定是正确的。[99]

后来，弗雷德尔与利维因（Etera Livine，来自法国大溪地，圆周研究所的博士后）合作，详细证明了DSR在有物质的2+1维引力理论中是如何成功的。[100]这些结果很重要，因为DSR有了一个可能世界的模型，也就保证了理论的和谐。

为了使DSR成为可靠的理论，还有一个必须解决的问题。我们说过，在许多形式的理论中，存在某个粒子所能具有的最大能量，通常被认为是普朗克能量。这不是一个实验问题，因为观测到的最大能量是在AGASA宇宙线探测器中的质子能量，大约是那个最大能量的十亿分之一。

不过乍看起来，这个能量限制似乎适用于任何物体：不仅电子或质子，包括狗儿、恒星和足球，都应该小于那个最大能量。这显然与自然矛盾，因为任何具有多于1019个质子的系统都有大于普朗克质量的能量。狗儿大约有1025个质子，恒星更多。这就是所谓的足球问题。

足球问题存在于二维世界，但没有必要去解决它，因为我们不在那个世界做实验。我们只需要知道，在那个世界里，任何物体，不论由多少粒子组成，其所有的能量都小于普朗克能量。

足球问题有一个自然的解决，大概在我们三维空间的世界里成立。若昂和我以前就提出过这个解。我们的观点是，物体所含的每个质子具有等于一个普朗克能量的最大能量。这样，包含大约1025个质子的足球的能量不可能高于1025个普朗克能量。那么也就不存在观测问题了。

我们可以看到这个解是成立的，但我们不知道它为什么一定正确。最近，利维因与圆周所的另一个博士后吉雷利（Florian Girelli，也来自法国）提出一个解释。他们发现了一种奇妙的方法来重构理论，那个解会从中自然冒出来。[101]既然足球问题解决了，我不知道还有什么能阻碍DSR成为我们世界的正确理论。奥格和GLAST在未来几年的观测大概也能证明它；如果不是，它至少也将被证明是错误的，这说明DS R是一个真正的科学性的理论。

我们现在可以回来看，假如狭义相对论失败了，对不同的量子引力理论有什么意义呢？我们已经看到，失败可能有两方面的意义，那要看实验告诉我们什么。狭义相对论可能在这个尺度上完全失败，意味着运动与静止确实存在着绝对的区别。或者，狭义相对论可以保留下来，但需要深化，如DSR。

弦理论在这两种改变下还成立吗？所有已知的弦理论肯定都将证明是错误的，因为它们强烈依赖于狭义相对论的成立。但会不会有某个形式的弦理论能与那两种失败之一相容呢？有几个弦理论家坚决地对我表示，即使狭义相对论会失败或需要修正，总有一天也可以构造一种能满足任何实验观测的弦理论。他们的意见可能是对的。我们说过，弦理论有许多没有观测到的场。有很多方式改变弦理论的背景，从而出现某个优越的静止状态，也就否定了运动的相对性。以这种方

式大概可以构造出一种能符合实验的弦理论。

那么DSR呢？会有与它相容的弦理论吗？我写本书时，若昂和我是仅有的两个思考这个问题的人，我们发现的证据很复杂。我们能构造一个能满足某些一致性检验的弦理论，但我们没能发现对其他检验的明确结果。

因此，尽管所有已知形式的弦理论都与狭义相对论相容，尽管狭义相对论可能有问题，但弦理论家还是可以协调这样的发现。令我疑惑的是，为什么弦理论家认为这有助于他们的事业。对我来说，它更加意味着弦理论不能做出任何预言，因为它只不过是理论的集合，而其中的每一个理论都对应着众多可能背景的一个。在GLAST和奥格观测实验中，关键问题是空间和时间的对称性。在一个背景相关的理论中，这取决于背景的选择。只要理论允许，通过选择恰当的背景，就能得到需要的答案。这和预言是截然不同的事情。

其他量子引力方法又如何呢？有什么方法预言过狭义相对论的失败吗？在背景独立的理论中，情形大为不同，因为时空几何不再决定于背景的选择。那个几何必须作为理论的一个解而出现。量子引力的背景相关方法必须做出真正的关于空间和时间对称性的预言。

我以前讲过，如果世界有两个空间维，我们是知道答案的。那种情形下没有自由选择，计算表明粒子运动遵从DSR。在有着三个空间维的真实世界里，也有同样的结果吗？我的直觉告诉我那是可能的，我们在圈量子引力中的结果为这个思想提供了证据，但还不算是它的证明。我最大的希望是，这个问题在实验观测告诉我们结果之前就可以很快解决。从引力的量子理论中得到一个真正的预言，然后让确凿的观测来证明它是错误的，那该是多么奇妙的事情啊！更好的事情当然是实验证明预言是正确的。不论哪种情形，我们都在做着真正的科学。

# 第十五章　弦论后的物理学

在前面两章我们看到，有理由期待我们对自然律的追寻将取得巨大进步。令人惊奇的实验发现已经拓展了相对论，它本身就为正在进行的实验提供了预言。不管双狭义相对论正确与否，它是真正的科学，因为正在进行的实验可以证明也可以否定它的主要预言。

理论家和实验家们（他们的工作在前两章讲过了）已经开辟了基础物理学的后弦论时代。在这一章，我将带领大家游历这个新世界，见识最有希望的思想和发展。在弦论以外，我们看到，通过对基本问题的艰苦而执著的思索，通过数学和物理学的同步发展，很多传统方法的基础物理学复苏了。在所有的前沿领域——量子引力、量子物理学的基础、基本粒子物理学和宇宙学，大胆的新思想接二连三地随着迷人的新实验产生出来。这些思想的萌芽一定会茁壮成长起来的，否则它们早就夭折了；它们展现了美好的前景。

我们先说一个正在进步的领域：一种量子引力方法，它没有逃避爱因斯坦的伟大发现，而是继承了它——即时空几何是动力学的、偶然的。

我曾多次强调，用空间摆动的弦来构造一个包含引力子的理论是不够的。我们需要一个关于空间由什么构成的理论，一个独立于背景的理论。前面讲过，广义相对论的成功说明空间几何不是固定的；它是动力学的，随时间演化的。这是不容颠倒的基本发现，任何未来的理论都必须包容它。弦理论没有，所以，如果弦理论成立，那么它的背后一定还有更基本的理论——一个背景独立的理论。换句话说，不论弦理论是否成立，我们都需要去寻求一个背景独立的量子引力理论。

幸运的是，通过最近20年的研究，我们对如何构造那样的理论有了很多认识。背景独立的量子引力方法研究是从1986年开始的，那时

第一次弦理论革命刚过去两年。催化剂是印度理论物理学家阿什特卡（Abhay Ashtekar，当时在锡拉丘兹大学）发表的文章。他重构了广义相对论，使方程更加简洁。[102]有趣的是，他把爱因斯坦的理论表达成接近规范理论的形式，而规范理论是粒子物理学标准模型的基础。

不幸的是，多数弦理论家并没注意过去20年在量子引力领域的这些进步，所以两个领域是独立发展的。外行人看这种互不往来的状况也许觉得奇怪。在我看来也很奇怪，所以我才想尽力去扭转它，向每个阵营宣传对方的长处。但我不能说取得了多大成绩。我之所以认为物理学正在危机中，部分原因就是我没能让用不同观点研究同样问题的人相互沟通——我也努力在思考如何改变这种局面。

量子引力领域的整个氛围都不同于弦理论。它没有宏大的理论，也没有时尚和潮流，只有几个很好的人辛勤地思考着几个密切相关的思想。我们有不同的研究方向，但也有某些统一的思想将这个领域连成一个和谐的整体。

主要的统一思想说起来很简单：不要从空间或任何在空间运动的东西开始。从纯粹的量子力学的而且具有纯量子结构的东西（而不是空间）开始。假如理论是正确的，那么空间会自然出现，代表那种结构的一般性质——犹如温度的出现代表原子的运动。

于是，很多量子引力理论家相信存在一个更深层的实在，那里不存在空间（这是背景独立性的逻辑极端）。因为弦理论需要背景独立理论的存在才有意义，许多弦理论家也表示他们认同这一点。在一定的意义上，如果强形式的马尔德希纳猜想（见第九章）是正确的，那么会从固定的三维几何里生出一个九维几何。难怪我们听威藤说，“借凝聚态物理学的名词来说，多数弦理论家猜测时空是一种‘突现现象’。”（他最近在圣巴巴拉加州大学卡夫利理论物理研究所的一次讲话。）[103]

终于有弦理论家开始欣赏这个观点了，我们就等着他们去研究已经获得的具体结果，继续前进。但在量子引力的大多数人的心里其实装着比马尔德希纳猜想更激烈的东西。

我们的出发点与几何毫不相干。量子引力学者们说空间是突现的，意思是空间连续体是一种幻觉。正如水面和丝绸的光滑隐藏了物质由离散的原子构成的事实，我们猜想空间的光滑也不是真实的，空间是作为某种基本材料（我们可以计数的）构成的某种东西的近似而突现出来的。在有的方法中，干脆假定空间是离散的“原子”组成的；在另一些方法中，这种假定是通过结合广义相对论与量子理论而严格推导出来的。

另一个统一的思想在于因果性的重要。在经典广义相对论中，时空几何决定了光线如何传播。因为没有比光更快的东西，一旦知道了光的传播，就能决定某个特殊事件引发了哪个事件。对两个发生的事件，只有当一个粒子以光速或更低的速度从甲传播到乙，我们才能说甲是乙的原因。因此，时空几何包含了一个事件引发另一个事件的信息。这就是时空的因果结构。

并不只是时空几何决定因果关系，也可以反过来：因果关系能决定时空几何，因为只要知道了光线如何传播，决定时空几何所需要的多数信息也就确定下来了。

空间或时空从某种更基本的东西突现出来，说起来很容易，但要发展这个思想，真正将它实现，却是非常困难的。实际上，以前尝试过的几种方法都失败了。我们现在明白，它们的失败是因为忽略了因果性在时空中的作用。如今，我们多数做量子引力的人相信，因果性本身才是基本的——即使在空间概念消失的地方它也有意义。104

目前最成功的量子引力方法结合了三个基本思想：空间是突现的，空间的最基本的描述是离散的，这些描述都以因果性为基本要

素。

当前的量子引力研究在某些方面很像100年前的物理学，那时人们相信原子，但不知道原子结构的细节。尽管不知道细节，玻尔兹曼、爱因斯坦等人还是仅凭物质由原子构成的事实认识了很多关于物质的东西。他们只知道原子的近似大小，却能预言可以观测的效应。同样，我们根据基于三个原则（突现、离散、因果）的简单模型，也能导出一些重要结果。在不知道细节的情况下，这些模型对空间的离散单位做了最简单的可能假定，然后看它们能产生什么东西。其中最成功的模型是罗尔（Renate Loll）和安比约恩（Jan AmbjØrn）构建的，叫因果动力学三角化。[105]这个名字也许太专业，而方法的思想却非常简单，就是用简单的构建材料来代表因果过程，看起来就像小孩子玩儿积木（图15-1）。也许该称它为富勒（Buckminster Fuller）方法。[106]基本思想是，时空几何由大量基本元素构成，每个元素代表一个简单的因果过程。这些元素如何堆砌，由几个简单的法则决定；每一个这样的量子时空模型的量子力学几率，则由一个简单的公式来计算。

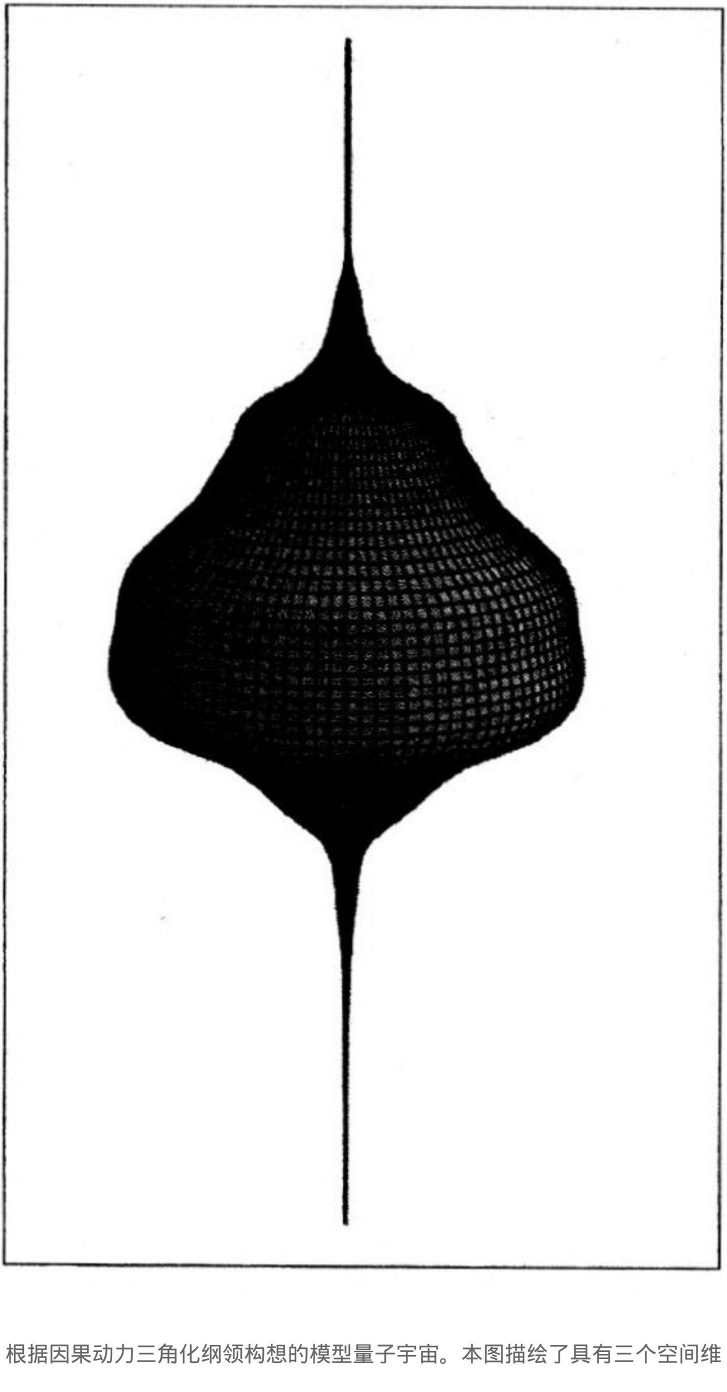


图15-1　根据因果动力三角化纲领构想的模型量子宇宙。本图描绘了具有三个空间维（其中一

维是水平的）和一个时间维（垂直方向）的模型量子宇宙的演化历史（蒙 Renate Loll 同意）

罗尔和安比约恩设定的一个法则是，每个量子时空都必须看作一个前后相接的可能空间的序列，犹如世界时钟的一个个瞬间。时间坐标是任意的，与广义相对论的一样；但不同的是，世界的历史不能再看作是一个在时间上前后相接的几何序列。

在这个限制下，再加几个简单法则，他们就得到重要的证据，说明经典时空（三维空间连同一维时间）从简单的积木游戏中突现出来了。如果说在背景独立的量子引力理论中，三个空间维的经典时空可以从仅基于离散性和因果性的纯量子世界突现出来，那么这至今还是最好的证据。特别是，安比约恩等人证明，如果不加关于因果性的约束，就不会有经典时空几何的出现。

因为这些结果，过去人们广泛相信的一些量子引力思想现在看来其实是错误的。例如，霍金等人常说因果结果是非基本的，在量子引力的计算中，可以忽略时间与空间的差别——即使在相对论中，这种区别也是存在的——而仅将时间看作另一维的空间。霍金在他的《时间简史》中所谓的神秘的“虚”时间就是这个意思。安比约恩和罗尔的结果表明这种想法是错误的。

在他们的工作之前，已经有人考察过时空的基本构成元素包含着因果性思想，但没想到经典时空可以突现出来的理论。有一种叫因果集合理论的形式，认为时空的基本单元都是裸事件，都可以归结为一系列引发它们和它们能引发的其他事件。这些思想比罗尔和安比约恩的模型还要简单，因为它们不需要整体的时间序列。现在还不可能让经典时空从这个理论突现出来。

不过，因果集合理论也有一个主要成果，那就是它似乎解决了宇宙学常数问题。锡拉丘兹大学的物理学家索金（Rafael D.Sorkin）和他的合作者们，通过简单假定经典世界来自因果集合理论，就预言宇宙学常数大致与观测的结果一样小。据我所知，这是迄今对宇宙学常

数问题的唯一明确的解答。这个解本身已经够吸引人了，何况理论的假定又是那么简洁，它当然成为一个值得继续探索的研究纲领。

英国数学物理学家彭罗斯（Roger Penrose）也提出一种量子时空方法，它所依赖的基本原理就是因果关系是真正基本的东西。他的方法叫扭量理论，他和几个追随者从20世纪60年代起就在做了。它的基础是将传统的观察时空事件的方法颠倒过来。人们总是习惯将发生的事件看成基本的，而将事件之间的关系看成第二位的。因此，事件是真实的而事件之间的因果关系不过是事件的属性。彭罗斯发现这种观察事物的方法可以颠倒过来。你可以将基础的因果过程作为基本的，然后用因果过程之间的重合来定义事件。更具体地说，你可以接着将所有物理移到光线的空间里。其结果是美妙无比的结构，彭罗斯称它为扭量空间。

扭量理论在彭罗斯提出的前20年里发展很快。许多物理学基本方程都能以令人惊讶的美妙的方式改写成扭量空间的形式。仿佛你真的可以把光线看作最基本的东西，而空间和时间不过是它们之间的某种关系。它也为统一带来了进步，因为描述不同粒子的方程在扭量空间里都有同样简单的形式。扭量理论部分实现了时空可以从其他结构中突现出来的思想。我们时空的事件不过是悬浮在扭量空间的一些特殊的曲面；我们时空的几何也可以从扭量空间的结构中突现出来。

但这幅图景存在着问题。主要问题在于，扭量空间只有在没有量子理论的时候才能理解。虽然扭量空间与时空有很大区别，但它仍然是光滑的几何结构。还没人知道量子扭量空间像什么样子。量子扭量理论是不是有意义，时空是不是能从它突现出来，目前都还是未知的。

扭量理论在20世纪70年代的中心是牛津，我和很多人一样，也曾被吸引到那儿去度过一段时光。我感觉那儿的氛围很令人振奋，不像后来在弦理论中心形成的那种氛围。彭罗斯令人敬慕，后来的威藤也

成了那样的人。我遇到过一些天才卓绝的年轻物理学家和数学家，他们都深信扭量理论。有几个已经成了杰出的数学家。

扭量理论当然给数学带来了重要进展。它使我们更深入理解了几个重要的物理学问题，包括杨-米尔斯理论的主要方程，那是粒子物理学标准模型的基础。扭量理论也让我们更好地认识了爱因斯坦广义相对论的某些解。这些认识充分表现在几个不同的发展，包括圈量子引力。

但扭量理论尚未成为可行的量子引力方法——主要因为它没有办法包容广义相对论。但彭罗斯和几个同事仍然没有放弃。威藤领导的几个弦理论家最近也开始做那方面的工作，为扭量空间带来了一些新的方法，进展很快。这个方法似乎不能帮助扭量理论演进为引力的量子理论，但它正在变革规范理论的研究——如果需要什么证据，这正好说明我们不应该那么长久地忽略扭量理论。

彭罗斯并不是唯一为自己创造量子引力方法的一流数学家。也许最伟大的——当然也是最有趣的——健在的数学家是阿兰·康尼斯（Alain Connes），他是马赛的一个警察的儿子，大半生都在巴黎。我喜欢和阿兰谈话。有时我不明白他说什么，但他深刻的思想和绝妙的笑话令我快乐无限。（那些笑话经常是少儿不宜的，尽管有时说的是黑洞或讨厌的卡丘流形。）有一次，他在一个会上讲量子宇宙学，要我们每次听到宇宙一词时都站起来，以表示尊敬，引得哄堂大笑。不过，虽然我不能总是理解阿兰，他却总能理解我。他属于那种思维敏捷的人，知道你想什么，当然也能更好地说出你想说的话。尽管他和他的思想都那么自在轻松，却一点儿也不爱争斗，对别人的思想总是怀着真诚的好奇。

阿兰的量子引力方法要回到基础，发明一种能完美统一几何的数学结构与量子引力的新数学。这种数学我在第十四章提起过，叫非对易几何。“非对易”指量子理论的量由不能对易的对象来代表，即AB不

等于BA。量子理论的这种非对易性密切关联着这样一个事实：不能同时测量一个粒子的位置和动量。当两个量不对易时，就不能同时知道它们的数值。现在，这似乎与几何的本质相矛盾，因为几何就是从曲面的直观图像出发的。而形成直观图像就意味着需要完整的定义和完全的知识。要把几何那样的东西建立在不能同时知道的事物基础上，其实是很重大的一步。它的诱人之处就在于它在使自身成为物理学下一步的恰当数学的同时，也以新的方式统一了几个不同的数学领域。

非对易几何出现在几种不同的量子引力方法中，包括弦理论、DSR和圈量子引力。但这些理论没有一个深层把握了康尼斯原先的概念，他和几个数学家（多数是法国的）还在继续发展着。[107]它出现在其他纲领中的不同形式都是从概念的表面出发，例如将空间和时间的坐标变成不对易的量。康尼斯的思想要深刻得多；它在根本上统一了代数和几何。只有像他那样，既探索数学又创造性、战略性地思考数学知识结构及其未来的人，才可能创造那样的思想。

和老的扭量理论家一样，康尼斯的追随者也是忠心耿耿的。宾州大学要举办一个关于不同量子引力方法的会议，阿兰举荐了一位年长的法国著名数学家，叫卡斯特勒（Daniel Kastler）。会前一个星期，这位先生骑车摔断了腿，但他爬出了医院，独自到了马赛机场，正好赶在开幕时到达会场，他宣称，“有一个真阿兰，我是他的信使。”看来，并不只是弦理论家才有真正的追随者，非对易几何学者也有，而且更幽默。

非对易几何的一个成功是它直接引出了粒子物理学的标准模型。正如阿兰和他的同事们所发现的，当你将麦克斯韦电磁学写成最简单的非对易几何形式时，统一电磁力与弱核力的温伯格-萨拉姆模型将自然出现。换句话说，弱相互作用连同希格斯场都将自动而正确地显现出来。

回想一下我们在第二章说的，判断一个统一是否成功，就看它是

否能立刻表现与自然的一致。康尼斯的简单思想出现了正确的弱力与电磁力的统一，这是很诱人的结果。弦理论本该出现这样的东西，可惜没有。

还有一套方法也是关注经典时空和粒子物理学如何从基本的离散结构产生出来。这些是凝聚态物理学家们创立的模型，如斯坦福大学的劳克林，赫尔辛基技术大学的沃洛维克和MIT的文小刚。最近，这些方法被年轻的量子引力学者们采纳了，如德雷尔。这些模型是粗糙的，但也的确说明狭义相对论的某些特征（如速度的普适上限）可以从一定的离散的量子系统突现出来。沃洛维克和德雷尔煽动性地宣扬说，宇宙学常数问题解决了——因为它本来就不是问题。他们声称那种说它有问题的想法是错误的，那是因为太看重背景相关理论的结果。他们指出，这种错误源自人们将理论的变量割裂开来，将其中的一些作为固定背景，而将另一些作为量子场。[108]如果他们的意见在这一点上是对的，那将是多年以来从量子引力产生的最重要的结果。

我这里描述的所有方法都是背景独立的。有几个一开始就假定时空由离散元素构成。有一个做得更好，证明空间和时间的离散性是结合量子理论与狭义相对论的结果。这就是圈量子引力的成果。它的出发点是阿什特卡1986年对爱因斯坦广义相对论的革命性重构。我们发现，不需要添加任何东西，仅仅用一组新变量来改写爱因斯坦理论，就可以精确导出一个量子时空来。

圈量子引力的基本思想其实很老了，我们在第七章已经讨论过。它源自直接以场线来描述场（如电磁场）的思想。（之所以叫“圈”，是因为在没有物质的情况下，场线可以自我闭合形成圈。）这是涅尔森、波里亚柯夫和威尔逊的观点，也正是这种思想引出了弦理论。弦理论基本上就是这种直观图景在固定时空背景下发展起来的。圈量子引力是同样的思想，然而是在完全背景独立的理论中发展的。

这个思想成为可能，全赖阿什特卡的一个重大发现：广义相对论

可以用规范场的语言来表述。这样，时空的度规成了电磁场一样的东西。当我们以量子力学方法处理对应的场线时，被迫不要背景，因为本来就没有——场线已经描述了空间的几何。一旦我们使它们成为量子力学的，就不会再留下经典的几何了。所以，为了摆脱背景度规，我们必须重新构建量子场论。长话短说，经过很多有着不同物理学和数学才能的人的努力，我们终于成功了。结果就是圈量子引力。

新图景很简单。量子几何是一种特殊的图（图15-2）。一个量子时空就是一个事件序列，其中的图在结构中因局部改变而演化。这最好用例子来说明，如图15-3。

这个理论带来了很多成功。它在以下三方面被证明是有限的：

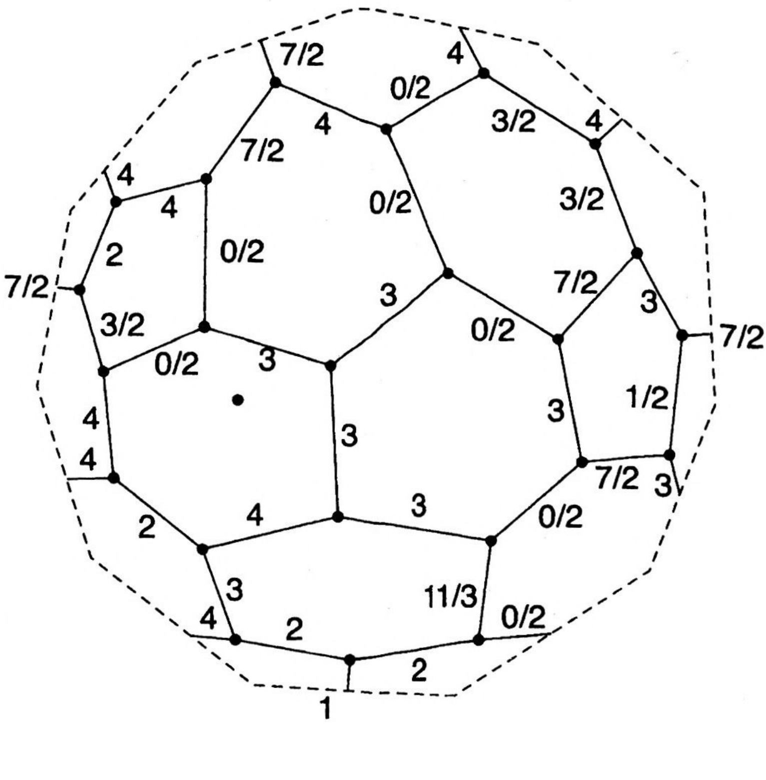



图15-2　自旋网络——圈量子引力及其相关理论中的量子几何状态。节点和边界分别联系着体

1.量子几何是有限的，从而面积和体积以离散单元形式出现。

2.计算量子几何演化为不同历史的几率时，结果总是有限的（至少在所谓巴勒特-克雷恩（Barrett-Crane）模型的特定形式下是有限的）。

3.当理论与物质理论（如粒子物理学标准模型）耦合时，通常出现的无限会成为有限。就是说，没有引力时，需需要强调的是，前面的陈述没有什么不确定性。圈量子引力的主要结果都经过了严格定理的证明。

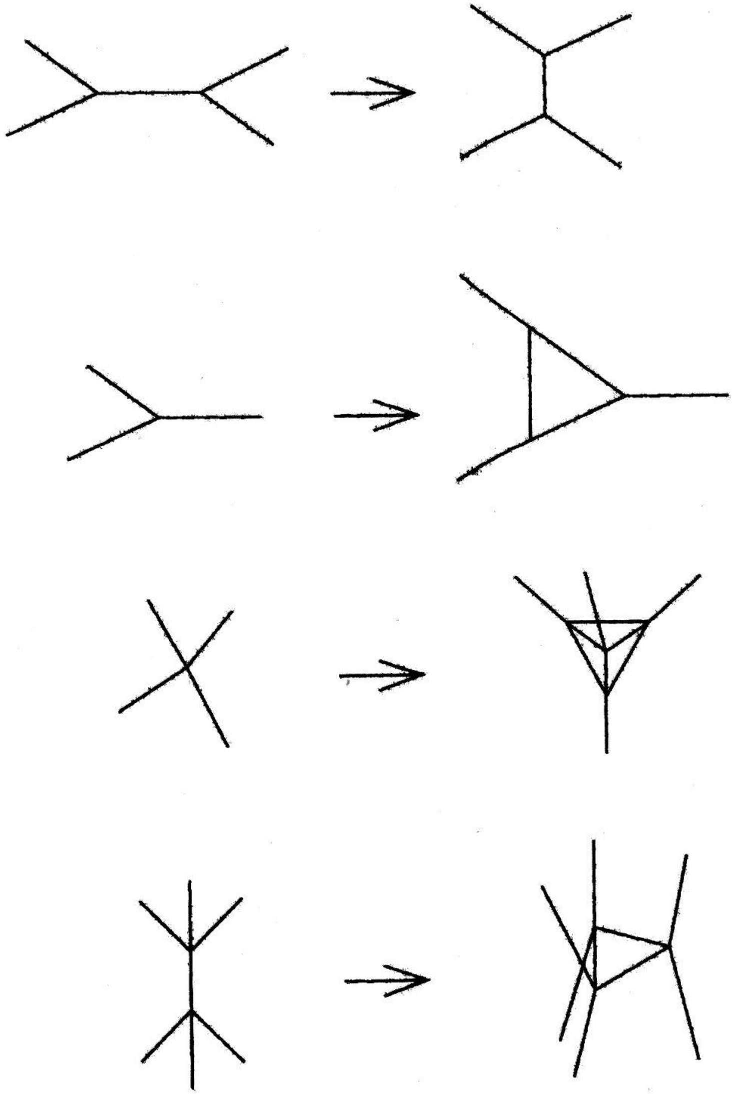


图15-3　自旋网络通过一系列这些局部改变而随时间演化

要用一个特殊过程将无限的表达式独立出来，使其成为不可观察的；有引力时，根本就不存在无限的表达式。

圈量子引力从一开始就面临的最大挑战是需要解释经典时空是如何实现出来的。最近几年，由于有了新的近似计算方法，这个问题取得了重大进展。这些新的过程证明理论具有描述宇宙的量子态，其几何在很好的近似下是经典的。去年，马赛理论物理中心的罗维利（Carlo Rovelli）和他的同事们迈出了重要的一步，他们第一次证明圈量子引力预言了两个物体会以牛顿定律所决定的方式相互吸引。[109] 这些结果也表明理论在低能量近似下有引力子，因此圈量子引力的确是一个引力理论。

人们如今正想着把圈量子引力用于现实世界的现象。它精确描述了黑洞视界，能得到正确的熵。这些结果与贝肯斯坦和霍金过去关于黑洞有熵和温度的预言（见第六章）是一致的。我写这些的时候，研究生和博士后们正在做的一个热点问题是修正霍金关于黑洞热力学的结果，假如能在未来的自然黑洞研究中观测到那些现象，那就能证实或证伪圈量子引力。

圈量子引力也是进一步研究黑洞内部随时间强烈变化的几何的基础。有几个计算表明，黑洞内的奇点是可以去除的。因此时间能在经典广义相对论预言的终结点外延续下去。它流向何方呢？似乎进入一个新生的时空区域。奇点就这样被我们所谓的时空反弹取代了。在反弹之前，黑洞内的物质在收缩，而反弹过后，它开始膨胀，进入一个以前不存在的新区域。这是非常令人满意的结果，它证明了德维特和惠勒以前的一个猜想。同样的技术还用于研究极早期宇宙发生的事情。理论家们也发现奇点同样被清除了，这意味着宇宙在大爆炸之前就存在了。

黑洞奇点的清除自然解决了霍金的黑洞信息疑难。如第六章说的，信息没有丢失，而是进入了新的时空区域。

根据圈量子引力对极早期宇宙的把握，我们可以为实际观测计算预言。圆周理论物理研究所的两个博士后霍夫曼（Stefan Hofmann）

和温克勒（Oliver Winkler）最近导出了量子引力效应的精确预言，有可能在未来的宇宙微波背景观测中看到。110

理论家们还忙着预言奥格和GLAST实验会看到什么，这两个实验将说明狭义相对论是否在普朗克能量失败。背景独立方法的一大好处是能为这样的实验提出预言。惯性坐标系的相对性原理是将继续保留还是被打破？或者像DSR理论那样修正？我已经强调过，没有一个背景相关理论能对这些实验做出真正的预言，因为问题已经通过背景的选择来回答了。特别是，弦理论认为惯性系的相对性还是正确的，而且和爱因斯坦在狭义相对论中的原始形式一样。只有背景独立方法能预言狭义相对论的这个原理的命运，因为经典时空的性质是以动力学问题的解的形式出现的。

圈量子引力承诺能做出明确的预言。在空间只有二维的模型中，它已经做到了：它预言DSR是正确的。有迹象表明，同样的预言在我们的三维世界也成立，不过目前还没有可信的证明。

其他大问题又如何呢？例如粒子和力的统一？直到最近，我们还认为圈量子引力对量子引力以外的问题几乎无话可说。我们可以在理论中加入物质，而好的结果不会改变。如果需要，我们可以将整个粒子物理学的标准模型或其他任何我们想研究的粒子物理学模型加进来，但我们并不认为圈量子引力对统一问题有任何具体明确的贡献。最近我们意识到我们错了。圈量子引力已经有了基本粒子，最近的结果表明那恰好就是那个正确的粒子物理学：标准模型。

去年，马科普洛提出了一种新的方法来解决空间几何如何从更基本理论自然出现的问题。马科普洛是做量子引力的年轻物理学家，经常提出一些几乎不可能的思想，结果却是正确的，令我惊讶不已，这是她最好的一个结果。她不直接问量子时空的几何是否能以经典时空的形式出现，而是基于确认和研究粒子在量子几何中的运动，提出了一个不同的方法。她的想法是，粒子必须是量子几何的某种突现的激

发态，它们像波穿过固体和液体那样穿过那种几何。然而，为了生成我们知道的物理学，这些突现的粒子只能作为纯量子粒子才能描述，而不管它们穿过的量子几何。[111]

通常情况下，如果粒子与环境相互作用，则粒子状态的信息会散失到环境中——我们说它脱散了。很难阻止脱散的发生；这也是量子计算机难做的原因，因为计算机的功效全赖粒子处于纯量子态。做量子计算机的人了解量子系统在什么时候（即使与环境接触）处于纯态。马科普洛在和那个领域的专家一起工作时，发现他们的认识也适用于说明量子粒子如何从量子时空生成。她指出这一点，是为了从量子引力理论得到预言，我们要做的只是识别那样的量子粒子，然后证明它像在寻常空间一样运动。在她的类比中，环境是动力学的、不断变化着的量子时空。量子粒子穿过它一定就像穿过固定的非动力学的背景一样。

不难说明许多背景独立的量子引力理论都有突现的满足马科普洛发现的那些条件的粒子。但那些粒子是什么呢？它们对应于我们见过的什么东西吗？

这个问题乍看起来很难，因为圈量子引力预言的量子几何非常复杂。粒子态联系着三维空间里的图。空间是背景，但除了本身的拓扑而外不具有其他性质；所有关于几何度量的信息——如长度、面积和体积——都来自那些图。但因为图必须画在空间，理论就有很多额外的信息包含在其中，而它们似乎与几何无关。这是因为图在三维空间里可以通过无限多种方式纽结、连接和缠绕。

图的纽结、连接和缠绕有什么意义呢？这个问题从1988年起就伴随着我们。我们一直不知道它们到底意味着什么。马科普洛的方法使我们更急迫想知道问题的答案。答案突然来了。

去年春，我偶然看到一个年轻的澳大利亚粒子物理学家的论文，

他的名字令人难忘，叫比尔森-汤普森（Sundance O.Bilson-Thompson）。他在文章里展示了一根简单的丝带辫子，准确把握了我在第五章讲过的粒子物理学的前子模型的结构特征。（回想一下，这些模型假定所谓的前子为质子、中子和标准模型认定的其他基本粒子的基本构成。）在他的模型里，前子就是丝带，不同类型的前子对应于沿不同方向缠绕或未缠绕的丝带。三根丝带可以缠绕在一起，不同的缠绕方式正好对应于标准模型的不同粒子。[112]

我一看这篇论文，就发觉它是我们需要的思想，因为比尔森-汤普森研究的缠绕都会出现在圈量子引力中。这意味着在量子时空里图的不同纽结和缠绕方式必然是不同类型的基本粒子。所以，圈量子引力不但是关于量子时空的——它已经包含了基本粒子物理学。假如我们能发现在理论中运行的比尔森-汤普森游戏，它恰好就该是基本粒子物理学。我问过马科普洛，他的缠绕是否就是她的相干激发态。我们请比尔森-汤普森合作，经过几次失败的开始后，我们才明白这一点其实一直都是成立的。依靠几个小假定，我们发现了一个前子模型，描述了一大类量子引力理论中的最简单的类粒子态。[113]

结果引出了很多问题，回答它们是我当前的主要目标。它是否好到能为即将在CERN巨型重子对撞机（LHC）做的实验给出明确的预言，现在还说不准。但有一点是清楚的。弦理论不再是能统一基本粒子的量子引力的唯一方法。马科普洛的结果意味着许多背景独立的量子引力理论都包含着以突现状态存在的基本粒子。而个别理论，如圈量子引力，不会引出众多可能理论形成的景观。相反，它似乎导致唯一的理论，要么与实验一致，要么不一致。最重要的是，它使我们不必像苏斯金等人宣扬的那样（第十一章），需要求助人存原理来修正科学方法。传统方法做的科学正在前进着。

坦率地说，物理学的五个基本问题确实有着不同的解决方法。除了弦理论而外，基本物理学领域在几个方向上迅速进步，其中（不仅仅）包括因果动力三角化和圈量子引力。正如任何健全的科学领域一

样，实验与数学之间存在着强烈的相互作用。虽然从事这些研究纲领的人（大约300个左右）不像弦理论那么多，还是有很多人在挑战科学前沿的基础问题。20世纪向前的几大步是更少的几个人迈出的。当科学要发生革命时，重要的是思想的质量，而不是信奉者的数量。

不过，我想说明的是，在这个新的后弦理论的氛围里，没有排斥弦理论研究本身的东西。正如我前面讲的，它所依赖的思想——场与弦的对偶性——同样也是圈量子引力的基础。将物理学带进当前危机的不是这个中心思想，而是它在背景相关的条件下的一种特殊的实现——就是那种条件，将它与一些危险的概念捆绑在一起了，如超对称和更高的空间维。不同的弦理论方法——更适应背景独立等基础性问题和量子引力的诸多问题的方法——竟然可以不是最终发展的一部分，这是没有理由的。但为了认识这一点，弦理论需要在开放的氛围下发展，将其作为若干思想之一，而不要预先对它最后的成功或失败抱任何假想。新物理学精神所不能忍受的是，不顾任何证据而假定某个思想一定会成功。

虽然量子引力理论家们都为今天的进步感到兴奋，他们还是强烈期待未来至少会带来几个惊喜。量子引力的人们不像两次超弦革命期间意气风发的弦理论家，他们很少相信自己已经将最终的理论把握在手了。我们意识到，量子引力的背景独立方法的成功是完成爱因斯坦革命必须迈出的一步。他们证明，可以找到一种和谐的统一量子理论和广义相对论的数学和概念的预言。这种预言是弦理论拿不出来的，凭着这个框架，我们可构建一个能全面解决我在第一章列举的五大问题的理论。但我们也很清楚，我们并没有胜券在握。即使有最近的成功，还没有哪个思想看起来像是真的。

我们重温物理学的历史，会看到一个突出的事情：当正确的理论最终被提出时，它会很快赢得胜利。少数真正的好的统一思想是以动人、简单和唯一的方式出现的，它们没有众多的选择和可以调节的特征。牛顿力学由三个简单定律确立；牛顿引力由一个带常数的公式决

定。狭义相对论是一出现就完整的。完全建立量子力学经历了25年，但它从一开始就与实验同步发展着。自1900年以来，学科的许多关键论文要么解释最新的实验结果，要么为即将完成的实验做出明确的预言，广义相对论也是同样的情况。

因此，所有成功的理论都会产生实验结果，它们很简单，能在几年内得到检验。这并不意味着理论可以精确求解——多数理论都不可能。但它的确说明物理洞察直接导致了新物理效应的预言。

不管你想怎么说弦理论、圈量子引力和其他方法，它们都还没有达到那个前沿。通常的借口是说那个尺度的实验还无法实现——但我们已经看到，事情不是这样的。因此一定还有其他原因。我相信我们还缺失某个基本的东西，我们还在做着错误的假定。那样的话，我们就需要将错误的假定找出来，用新的思想来取代它。

那错误的假定会是什么呢？我猜它涉及两个因素：量子力学的基础和时间的本质。我们已经讨论过第一个；我发现，新的量子力学思想很可能最近已经在量子引力研究的驱动下提出来了。但我强烈感到关键问题还在于时间。我越来越觉得量子理论和广义相对论在深层上都把时间的本质弄错了。只结合它们是不够的。还有一个更深层的问题，也许要追溯到物理学的起源。

17世纪初，笛卡儿和伽利略做出了最奇异的发现：你可以画一张图，用一个轴做空间，另一个轴做时间。于是，穿过空间的运动成为图上的一条曲线（图15-4）。在这种方式下，时间仿佛成了另一维空间。运动被冻结了，匀速运动和变化的整个历史呈现在我们面前就像是静止不变的。硬要我猜想的话（猜想也是我谋生的手段），我想罪魁就在这儿。

我们需要寻求解冻时间的方法——不用将时间转化为空间来表示。我不知道该怎么做。我想象不出有什么数学可以不把世界表现为

永久冻结的样子。表示时间非常困难，因而时间的表示很可能就是我们失去的一环。

有一点很清楚：在弦理论的约束下思考这种问题，是不可能有什么结果的。因为弦理论仅限于描述在固定背景的时空几何中运动的弦和膜，对那些想开辟新天地、考虑时间和量子理论本质的人，它是无话可说的。背景独立方法提供了一个更好的起点，因为它们已经超越了经典的空间和时间的图景。而且，它们很好定义，很容易把握。还有一点好处是，它所用的数学很接近有些数学家已经用于探索时间本质的数学——即所谓拓扑斯（topos）理论的逻辑方法。

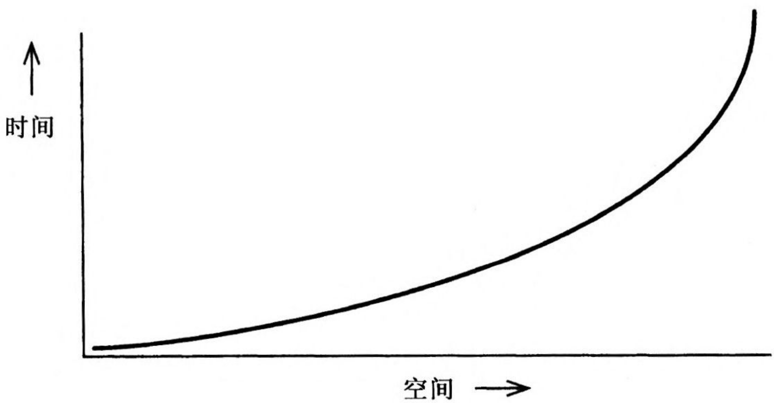



图15-4　笛卡尔和伽利略以来，在时间中展开的过程被表现为以另一维代表时间的图中的一条曲线。时间的这种“空间化”很有用，但可能的问题是它代表了一个静止的、没有变化的世界——一个冻结的永恒的数学关系的集合

关于如何表示时间而不将其转化为空间的一维，我明白一点，那就是它也出现在其他领域，从理论生物学到计算机科学，甚至还有法学。为了激发新的思想，哲学家翁格尔（Roberto Mangabeira Unger）和我在圆周理论物理研究所组织了一个小规模的学术讨论会，把相关领域的思想家们都请来讨论时间问题。那两天是我这几年最兴奋的日子。114

关于这一点我不想多说了，因为我想接着讲一个不同的问题。假定有一个野心勃勃的年轻人，他有着创新的头脑和坚韧，想深入思考那五个大问题。看到我们在那些问题遭遇的彻底失败，我不能想象他那样的人会甘愿陷入眼下的那些研究纲领。显然，如果弦理论或圈量子引力本身就是答案，那我们现在就该知道了。它们也许是起点，也许是答案的部分，也许包含着必要的教训。但正确的理论必须包含新的元素，也许只有我们雄心勃勃的青年才有能力去寻找它们。

我们这一代为年轻的科学家们留下了什么呢？思想和方法，他们大概不想用了；还有不同方向的局部成功的故事，结果却是未能完成爱因斯坦百年前开创的事业。最坏的情形是，我们将他们拉回来，要求他们继承我们的思想。于是，本书最后部分的问题是我每天早晨都问自己的问题：我们是不是尽力支持和鼓励年轻科学家——也包括我们自己——要超越我们在过去30年做的事情，去追求真正能回答物理学五大问题的理论？

# 第四篇　向经验学习

## 第十六章　奈何社会学

在本书最后一部分里，我想回到在引言里提出的问题。为什么经过了千百个才华卓绝而训练有素的科学家的那么多的努力，基础物理学在最近25年仍然没有什么确定的进步？面对那些前景在望的新方向，我们要做些什么才能确保前进的步伐能赶上1980年以前的200年？

可以从一个侧面来说明物理学的这种困惑，那就是在过去30年里，没有一个基本粒子物理学的工作能稳赢诺贝尔奖。原因在于，获奖条件是研究进展已经得到了实验的检验。当然，像超对称或膜世界那样的概念是可以通过实验来证实的，如果那样，它们的创立者应该得诺贝尔奖。但我们现在还不能说粒子物理学标准模型以外的哪个物理学假说的发现一定能获奖。

这种状况和我1976年刚读研究生时大不一样。那时人们很清楚，两年前才完成最后形式的标准模型是确实的进步。它已经得到了足够的实验证实，而且将得到更多的证实。它的发现者迟早会得诺贝尔奖，几乎是毫无疑问的。事实上，他们很快就得到了。

现在的情形完全不同了。过去25年有很多奖授予了理论粒子物理学的研究，但没有诺贝尔。诺贝尔不奖励聪明或成功，而是奖励正确。

这不是要否认这些研究计划取得了重大的技术进步。据说如今从

事研究的科学家比整个科学史上的人都多。物理学当然也是如此；今天大规模大学里的物理学教授比100年前整个欧洲（几乎所有科学进步的源地）的人数还多，所有这些人都在工作着，多数工作从专业上讲都是非常精深的。而且，今天年轻理论物理学家的专业水平都远远超过了他们的父辈和祖辈。年轻人有更多的东西需要把握，而他们竟然都做到了。

不过，假如我们以1980年前200年的标准来判断，基本粒子物理学的前进步伐似乎真的慢下来了。

我们已经讨论了最近25年失败的一些容易想到的原因。不是因为缺少数据，有足够的新结果能激发理论家的想象力；不是因为检验理论需要很长时间，从新理论预言新现象到实验证实很少有超过10年的；不是因为努力不够，今天做基础物理学研究的人远远超过了整个物理学史的人数。当然也不能抱怨说那些人缺乏才能。

在前面几章我曾假定，与其说失败是某个特殊的理论，倒不如说是一种研究风格。如果谁在弦理论家的营地和做背景独立的量子引力的人们中间游走一段时间，他一定会为两派人物研究风格和价值观念的巨大差异感到震惊。这些差别反映了理论物理学从半个世纪以前开始的分裂。

量子引力世界的风格继承了所谓相对论派的传统。这是爱因斯坦的助理以及他们的学生们所引领的——如贝格曼、戈尔博格（John Goldberg）和惠勒。这个群体的核心价值是尊重个人思想和研究纲领，怀疑流行时尚，依赖数学的纯净论证，他们深信关键的问题密切联系着关乎空间、时间和量子本性的基本问题。

另一方面，弦理论家群体的风格是基本粒子物理学文化的延续。这种风格总是充满着冲动、好斗和竞争的意气，理论家们喜欢争先响应新的进展（1980年前，那通常是实验的）而不相信哲学问题。当科

学的中心从欧洲转移到美国，当智力的焦点从基础新理论的解释转向理论的应用时，这种风格也取代了爱因斯坦和最子理论创立者们的那种思辩和哲学式的风格，而且成功了。

为了解决不同类型的问题，科学需要不同的风格。我的假定是，弦理论的错误在于它用基本粒子物理学的研究风格来发展，而那不适于新理论框架的发现。如果脱离实验，标准模型成功的风格也很难坚持下去。这种争斗的、赶潮流的风格只有在实验发现的武装下才能发挥作用，但在没有潮流而只有几个杰出人物的观点和品味时，它就只能失败。

我从20世纪70年代开始学物理时，两种研究风格都很健全。基本粒子物理学家比相对论专家的数量多，但两家都有活动的自由。对那些想解决空间、时间和量子理论的基本问题的人来说，虽然没有多少发展自己思想的天地，但只要有好的想法，也能得到足够的支持。从那时起，虽然相对论风格的需要增多了，它们在学术界的地盘却缩小了，原因是弦理论和其他重大研究计划成了主角。除了宾州大学的一个研究小组而外，做量子引力而不以弦理论或高维理论为基础的助理教授们，没有一个被美国研究大学录用的。

为什么最不适合眼前问题的风格却在国内外成了物理学的主角呢？这是一个社会学问题，但只要我们想构建思想，让我们的学科回到从前的成功，就必须回答这个问题。

为了说明问题的背景，我们看看科学景观的一些流行变化——年轻人若想开拓自己的科学生涯，就必须直面这片景观。

最显著的变化是，年轻人为了赢得有影响的老科学家的青睐，必须承受更大的压力。塑造美国科学辉煌的那一代人都接近退休了，他们也曾竞争过一流大学和研究机构的好职位，但是如果只想在某个地方做一个教授，有自己工作的自由，就不会有很多的压力。从20世纪

40～70年代，大学数量指数式地增长，年轻的科学家毕业时往往拥有几个大学的职位。我曾见过几个老同事，他们就从来没为找工作发过愁。

现在情况不同了。大学数量在20世纪70年代末就没再增加，而早先聘任的教授培养的研究生越来越多，这意味着物理学和其他科学的博士严重过量了。结果，研究型大学和学院的各个层次的学术职位的竞争都很激烈。学校更重视聘请那些可能得到研究机构的基金资助的人员。这就极大限制了年轻人发展自己的研究纲领的自由，他们只好跟着老科学家们的思路走。如果有创造力的人想找一个地方安静地做研究，追求冒险的原创的思想，那种地方是越来越少了。

与此相关的是，现在的大学比一两代人之前越来越专业化了。虽然大学教授的人数没有增加，管理者的人数和权利却增大了。因此，在招聘新人时，更少依赖个别教授的判断，而更多根据成绩的统计，如资助情况和论文水平。这也使年轻科学家更难逆着科学主流去开拓自己的新研究纲领。

为了公正地评价同行的工作，我们这些教授几乎总是反射性地党同伐异。即使我们上升到学术政治的高度，也常常免不了用单一的标准来评价一个科学家伙伴。在教授委员会和非正规讨论中，我们说某人“好”，某人“不好”，其实并不真的了解他们。个人一生的工作岂能简化为一句“老甲不如老乙好”？通常情况下，人们似乎更看重只靠聪明和勤奋就能取得的成绩，而不在乎探索性的思想或想象力。更重要的还是学界的风尚，谁要是忽略了它，就将断送学术生涯。

我曾和一个退休将军做项目，他曾领导过一个军官学院，后来成为一名商务顾问。他说与大学合作很伤脑筋。我问他问题出在哪儿，他说：“我们要向海军军官讲一个简单但基本的事情，那就是我遇到的大学管理者似乎没有一个知道，管理与领导是有重大区别的。后勤供给你只要管理就可以了，但在战场上你必须领导士兵。”我同意他

的话。我读大学时，见过的管理比领导多得多。

问题当然不限于科学。课程计划和教学方法的改革步伐确实还很落后。任何改革的建议都需要经过全体教授的批准，而多数教授一般都看不到他们几十年的教学方法有什么错误。我以前听说过顽固的大学怎么改革。我很幸运走进了一个第一年就学粒子物理学的学院。这是很难得的。尽管量子物理学在80年前就取代了牛顿力学，北美的多数院校却仍然把量子力学推迟到第三学年，而且只给物理学专业的学生上课。因为我知道怎么讲一年级的量子力学，我在哈佛大学读研究生时就建议那么做。我得到了年轻教授格奥尔基（Howard Georgi）的支持，同意和我一起讲课。但课本由人文与科学系的主任决定。他告诉我，这跟我们的设想无关，只是因为它没有经过那些不能逾越的委员会。“假如我们让每个教授都教他想教的，”他说，“我们的教育就会一片混沌。”我不知道教育混沌是不是那么可怕；不管怎么说，哈佛大学还是没有为一年级学生开量子力学课。

不幸的是，在美国拿物理学位的毕业生人数几十年来一直在减少。你大概认为这将缓解物理学职位的竞争。不是那样的，虽然本科生减少了，从发展中国家来的雄心勃勃的聪明博士却多了。同样的状况在其他发达国家也有。

我曾是耶鲁大学负责调查这种现象的教授委员会的一员（后来也就不管了），有机会问一些离开物理学的同学，他们为什么那么做。他们提出的一个理由是，物理学课程令人生厌——第一年不过重复他们在高中学过的东西，没有一点儿像量子理论、宇宙学、黑洞等令人兴奋的题目。为了稳住物理学生的下滑，我曾向聘我的每所大学建议把量子力学作为一年级的课程。每次我都被拒绝了，不过有两次允许我做小规模的量子理论课程试点。试点是成功的，听过课的少数学生现在已经有了很好的事业。

我并不是在这儿讨论课程改革，但这个例子说明大学并未很好地

起着创新机器的作用，哪怕只是革新落后科学80年的课程，他们也不敢冒一丝风险。

所有的科学家都悲叹各自领域的前进步伐。我认识几个生物学家和实验物理学家，他们痛苦地抱怨好多机会被浪费了，因为他们系里当权的老科学家们已经丧失了当年新做博士时的胆略和想象力。从学术界的底层提出的好思想得不到足够的重视；相反，高层人物的思想却被抬得太高了。

如果不考察助长这些风气的社会学，我们是不可能解决它们的。如果说我们物理学家自以为能解释基本的自然律，那么当然也该能理性地思考学术研究的社会学和那些令学术机构苦恼的阻碍科学进步的决策。

值得注意的是，“社会学”一词在弦理论家中间出现的频率高于我熟悉的任何其他科学家群体。它似乎是“群体观点”的简写。和年轻的弦理论家讨论问题的现状时，常听他们说“我相信这个理论，但我讨厌社会学。”如果你批评弦理论会议上报告的观点太局限，或者抱怨研究课题像走马灯一样变换太快，弦理论家会同意你的意见，而且补充说，“我不喜欢它，但那不过是社会学。”不止一个朋友劝告我，“群体已经决定了弦理论是正确的，你已经无事可做了。你不可能斗过社会学。”

真正的社会学家会告诉你，为了理解群体的作为，你必须考察权力。准对谁有权力？如何发挥那种权力？科学的社会学不是什么神秘的力量，它指的是成名的老科学家对青年科学家事业的影响。我们科学家说起它就感到不舒服，因为它迫使我们面对这样的事实：科学不全是客观和理性的。

但经过长时间的思考，我相信我们还是不得不谈理论物理学的社会学，因为我们笼统称之为“社会学”的那些现象对它的进步起着很大

的副作用。即使多数弦理论家都是真诚的人，怀着美好的科学愿望，但仍然存在一些社会学因素偏离了建立更大的科学共同体的理想。这已经在理论物理学方法中滋生了阻碍进步的病态。这个问题不在于弦理论是否值得做，值得支持，而在于为什么没有实验预言的弦理论能垄断基础物理学的资源，阻碍其他同样有前途的方法。有很好的证据说明，弦理论本身的进步也减缓了脚步，因为社会学限制了它要探索的问题，将科学进步所需要的具有想象力和独立思想的科学家挡在了门外。

应该指出，理论物理学中总是存在主导的领域。有段时间它是核物理，接着是基本粒子物理学。弦理论是最近的唯一的例子。任何时候都有一个主导的领域，也许物理学群体就是靠这种方式组织起来的。如果真是那样，我们需要考察为什么会这样。

弦理论群体首先令外人关注的就是他们的自信。作为1984年第一次超弦革命的见证人，我还记得当年欢呼新理论胜利的感觉。“未来12个月或18个月，一切都会好的。”弦理论的新星弗里丹劝我，“趁理论物理学还有事可做，你最好赶紧进来。”这只不过是一个例子，还有很多人都断言事情很快就会了结。

事情当然没有了结。但是，尽管后来经历那么多起伏，许多弦理论家仍然非常自信，既相信弦理论的真实，也相信自己比不能或不愿做弦理论的人更高明。对许多弦理论家（尤其是对过去的物理学没有记忆的年轻人）来说，简直不能理解一个天才的物理学家在面对如此良机时会选择做弦理论以外的事情。

这种态度当然伤害了其他领域的物理学家。斯坦福直线加速器中心的粒子物理学家赫维特（JoAnne Hewett）在她的博客里表达了自己的思想：

我发现有些弦理论家好狂妄，即使物理学家也没有那样的。有人

真的相信所有非弦理论家都低他们一等。他们相互的推荐信都那么写，我也听人当面那么说……弦理论[看起来]是那么重要，似乎所有其他理论都该为它牺牲。这有两种表现：弦理论家们占据了太多的教授位置，有时根本和他们的能力不相称，而年轻的弦理论家通常没有受过良好的实验物理学的训练。有些人简直连基本粒子的名字都叫不上来。这两种表现都是我们领域的远忧。115

赫维特博士描述的狂妄从一开始就是弦理论家群体的特征。钱德拉赛卡（Subrahmanyan Chandrasekhar，也许是20世纪最伟大的天体物理学家）喜欢讲一个故事。他80年代访问普林斯顿，人们为他获得诺贝尔奖举行庆祝会。在宴会上，他发现自己身旁坐着一个热心的年轻人。他以物理学家通常的方式开始了谈话，问他的邻座，“最近你做什么？”年轻人回答，“我在做弦理论，20世纪最重要的物理学进展。”接着，年轻的弦理沦家建议老钱放弃正在做的事情，改做弦理沦，否则他会像20年代那些没有紧紧抓住量子理沦的人一样落伍。

“年轻人，”老钱回答说，“我认识海森伯。我向你保证，海森伯决不会这么冒失地要人家停下正在做的事情而去做量子理沦。他当然也决不会无礼地向一个50年前就是博士的老人说他就要落伍了。”

任何与弦理论家打交道的人对这种超人的自信已经司空见惯了。不论讨论什么问题，从来没人（除非是门外汉）提出理论可能根本就是错的。如果提起弦理论实际上预言了一个景观而没有任何预言，有的弦理论家会大谈如何改变科学的定义。

有些弦理论家宁愿相信弦理论太深奥而不被大众理解，也不愿考虑它可能真的错了。最近某个物理学博客上有个帖子，很好说明了这一点：“我们不能指望一只狗理解量子力学，我们也许正在接近人类对弦理论认识的极限。也许天外还有先进的文明，在他们眼里我们也不过像一群小狗，也许他们已经认清了弦理论，而且发展了更好的理论……”116

其实，弦理论家似乎毫不怀疑弦理论必然是正确的，虽然他们也承认不知道它究竟是什么。换句话说，弦理论将包容后来的任何东西。第一次听说这个观点时，我以为是个笑话，但再三的重复令我相信说话者是认真的。赛伯格（Even Nathan Seiberg，普林斯顿高等研究院的著名理论家）在最近一次访谈中说（带着微笑），“如果[除了弦理论]还能有什么，我们会叫它弦理论。”[117]

他们还有一个相关的特征，就是自我感觉良好，而对用其他方法研究弦理论声明要解决的问题的人，他们却没有一点尊重。实际上，弦理论家对弦论之外的东西通常并不感兴趣，也视而不见。和量子引力会议不同，弦理论会议从不邀请其他路线的科学家们提交论文。这当然只会强化弦理论家们的断言：弦理论是唯一给出正确量子引力结果的方法。弦理论家不尊重其他方法，有时简直到了蔑视的地步。在最近的一个弦理论会议上，剑桥大学出版社的一个编辑私下告诉我，有个弦理论家告诉他绝不考虑让他们出版他的东西，因为剑桥出版了一本关于量子引力的书。这种事情并不罕见。

弦理论家知道他们在物理学世界的主导地位，多数都认为那是理所当然的——如果说理论本身还不能证明，有那么多天才人物为它工作这个事实也足以证明了。如果你向专家针对弦理论的某个论断提出具体的问题，那么你大概自己都没明白，就会被人家看作不可理喻，不是一路人。当然，更开明的弦理论家并不是这样的——但是，当年轻的弦理沦家突然发觉自己正在和一个并不欣赏他的假定的人谈话时，你常常会看到一张拉长的脸，我已经见得太多了，想不见都不行。[118]

弦理论的另一特征是，弦理论家与非弦理论家之间存在着明确的界线，这是其他物理学领域所没有的。也许你写过几篇弦理论的论文，但那并不意味着你一定会被弦理论家们认同为他们的一分子。起初我感到很疑惑。我曾走过一条老路，尽可能地向不同的方法学习。我最初也看到了自己做的事情，甚至包括量子引力，原来就是为了解

决弦理沦的一个未解问题——使它成为背景独立的形式。最后，有朋友告诉我，要想被弦理论群体认同——这样才有望做出成绩——你不但只能做弦理论，而且要做弦理论家们正在做的特殊问题。我想我的朋友大概没意识到那么做会伤害我的判断力，侵犯我的学术自由。

我兴趣很广，常参加我的领域外的会议。但只有在弦理论会议上，才会有人过来问，“你现在做什么？”如果我解释说我在做弦理论，想来看看其他人在做什么，他们常常会古怪地皱着眉头说，“你不是那个做圈的家伙吗？”在天体物理学、宇宙学、生物物理学或后现代主义的会议上，都不曾有人问我去那儿做什么。在一次弦理论会议上，一个著名弦理论家坐下来，伸出手说，“欢迎回家！”另一个人说，“很高兴在这儿看到你！我们一直为你担心呢。”

任何一年里，弦理论深入探讨的领域都不会超过两个或三个。研究领域逐年改变，只要看看人们在重要年会上谈论的题目，就能跟踪这股潮流。经常是至少三分之二的话题集中在一两个方向，它们两年前还在幕后，而过两年就将退出舞台。年轻人很清楚，成功的事业需要紧跟一两股时尚的潮流，正好可以赢得一个好的博士后位置，然后做一个好的助理教授。如果和弦理论的领导者们谈这个问题（我就谈过很多次），你会发现他们真的相信集中众多聪明人一起努力，会比让个人独立思考和探索不同的方向，能更快取得进步。

这种整体的和“有纪律的”（某个老资格弦理论家说的）方法产生了三个不幸的后果。首先，如果问题不能在两三年内解决，就会被扔掉，而且再也不会被捡回来。原因简单而残酷：年轻的弦理论家如果不快点儿放弃他们辛苦得来的落伍专业而走进新的方向，就可能丢掉饭碗。第二，整个领域仍然处在几个老人的思想和研究纲领的驱动下。在过去的10年里，只有两个年轻弦理论家，马尔德希纳和布索（Raphael Bousso），取得了改变领域方向的发现。这和更健全的物理学领域是截然不同的，那些领域的多数新思想和新方向都来自人们二三十年的研究。第三，弦理论没有充分发挥群体内的人才的作用。

有很多努力是重复的，而很多有潜力的重要思想却没人探索。任何身居大学委员会负责选拔博士后的人，都会看到路越走越窄了。在诸如宇宙学、量子信息论或量子引力等领域，有多少候选者就有多少个研究方案，而且很多思想是闻所未闻的。在弦理论的领域，你经常会一次又一次地遇到相同的两三个研究计划。

当然，年轻人知道自己在做什么。我在那样的委员会待过多年，发现弦理论家使用的能力评价标准几乎都不同于其他领域，只有很少的例外。据我所知，对当前问题所需要的数学能力的要求比原创性思想更高。如果某人只和一流科学家一起发表文章，而其研究计划表现不出独立的判断力和创造性，他是不会被一流的量子引力研究机构录用的，但这似乎是进入一流弦理论研究中心做博士后的最保险的途径。令我兴奋的申请者——有独立完成的论文，描述了令人惊奇的大胆的思想——却令弦理论的朋友们感到寒心。

在我经历过的其他群体，如量子引力和宇宙学，未解的问题都存在多种观点。如果你和五个不同的专家交谈，不论老的还是年轻的，你会得到五种不同的关于课题前景的认识。除了最近关于理论景观和人存原理的讨论而外，弦理论家有着惊人一致的观点。你可以从不同的人那儿听到同样的事情，甚至同样的言论。

我认识一些年轻的弦理论家，他们不承认有这些特征。他们坚持说群体内有多样的观点——只是旁观者不能走近罢了。这是好消息，但人们私下对朋友说的却不是这样的。实际上，如果在私下里而不是公开传播不同意见，就说明有种等级在决定着对话——以及研究过程。

弦理沦的领导者们对研究活动的打压，不仅应在原则上遭到谴责，它几乎已经肯定阻碍了进步。我们这么说，是因为我们看到领域内的很多思想要在第一次提出多年以后才变得重要。例如，弦理论是由大量理论组成的一个集合，这个发现第一次是斯特罗明戈在1986年

发表的，但到2003年，弦理论家才跟着卡洛什（Renata　Kallosh）和她在斯坦福的同事们的工作，开始广泛讨论它。[119]下面是CERN著名弦理论家莱尔歇（Wolfgang Lerche）最近的一段话：

我感到恼火的是，这些思想在20世纪80年代中期就出现了，在一篇关于四维弦构造的论文里，曾粗略地估计弦真空的数量，大约为101500。这项研究被忽略了（因为它不符合当时的哲学），而忽略它的人现在却重新“发现”了那片景观，并因此走进杂志，甚至还想写专著……整个讨论本可以（其实也应该）发生在1986年或1987年。那时以来，真正发生变化的是某些人的头脑，你现在看到的就是全力运转的斯坦福的宣传机器。[120]

我本人提出应将弦理论视为一个理论景观，最早发表在1992年，也没人理睬。[121]这并不是孤立的例子。其实，早在1984年第一次弦理论革命之前就出现过两个十一维超对称理论，但直到10多年后的第二次革命它们才被重新发现。它们分别是十一维超引力和十一维超膜。在1984年和1995年间，少数理论家构造了这些理论，但被推到了弦理沦群体的边缘。我还记得几个美国弦理论家可笑地谈论过这些“欧洲超引力狂”。1995年后，人们猜想这些理论应该与弦理论统一于M理论，而且欢迎过去做那些理论的人回到弦理论阵营中来。显然，如果这些思想没有被长久地忽略，进步会快得多。

有几个思想也许有助于弦理论解决它自身的问题，但它们还没得到广泛的研究。其中一个是老思想，即所谓的八元数系，是深入理解超对称与高维之间的关系的关键。[122]另一个是我已经强调过的，即要求弦理论或尚未探索的M理论的基本形式必须是背景独立的。在2005年弦理论年会关于“下一次超弦革命”的小组讨论中，斯坦福理论物理研究所所长申克尔指出，它很可能会在弦论之外的某个课题发生。如果说这是弦论头面人物的认识，那他们为什么不鼓励年轻人探索更广泛的课题呢？

弦理论的研究活动那么狭窄，也许是因为弦理论群体太尊重个别人的观点。在我遇到的科学家中，只有弦理论家在表达自己的意见之前会关心领域里的头面人物（如威藤）是怎么想的。当然，威藤思想清晰而深刻，但问题在于，如果过分把某个人的观点当成权威，对任何领域都不是好事。没有哪个科学家（包括牛顿和爱因斯坦）不曾在他们固执坚持的很多问题上犯错误。有很多次，在大会报告或谈话后的讨论中，如果出现了有争议的问题，总有人问，“那么，老威是怎么想的呢？”这常令我感到困惑，有时我会站出来说，“好的，如果我想知道老威怎么想，我会去问他；现在我问你怎么想，因为我对你的意见感兴趣。”

作对易几何就是一个例子，在威藤接受它之前一直被弦理论家忽略了。它的创立者康尼斯讲过下面的故事：

1996年我去芝加哥大学给物理系做报告。一个著名物理学家也在场，可讲话没完他就离开了。两年后，我在牛津附近的卢瑟福实验室的狄拉克论坛做相同的报告，又遇见他。他也来了，这次看起来很开明，好像是相信我讲的了。后来他讲话，很正面地提到了我的东西。这令我感到惊讶，因为那是同一个报告，而我没忘记他上次的反应。在回牛津的车上，我坐他旁边，问他，“你在芝加哥听我讲话时，怎么没听完就离开了？现在你真的喜欢它了吗？”那伙计不是初学者——已经40多岁了，可他的回答是，“我看见威藤在普林斯顿的图书馆读你的书！”[123]

应该说，这种态度少见了，也许是因为当前围绕那个景观有着太多的喧嚣。直到去年我才第一次听到来自弦理论家的怀疑。现在我能时常听年轻人说弦理论存在“危机”。“我们失去了领导者，”有人会说，“在这以前，热门的研究方向是什么，我们该做什么事情，总是很清楚的。现在没有真正的指南了。”或者，他们会神经质地问，“威藤真的不做弦理论了吗？”

弦理论令很多人感到恼火的另一方面是，它的实践者们，特别是年轻人，似乎都以救世主自居。对他们中的一些人来说，弦理论已经成了一个宗教。如果我们发表了质疑或批评弦理论家结果的文章，通常会收到的最友好的邮件大概是，“你在开玩笑吧？”或“这可笑吗？”很多网站和留言板都有弦理论的“反对者”们的讨论，即使在这样自由的场合，也有人用不堪入耳的言语质问非弦理论家的智力和专业能力。我们难免得出这样的结论：至少某些弦理论家已经开始将他们自己看作十字军战士而不是科学家了。

弦理论家们怀着那股傲慢，也用最乐观的眼光解读他们的证据。我的量子引力的同事对解决问题的前景一般都抱务实甚至悲观的态度。在圈量子引力理论家中，我大概算最乐观的，但和多数弦理论家比起来，也就暗淡无光了。面对重大的未解问题时，更是如此。前面说过，以“弦”的观点来看万物，是基于长期以来弦理论家们普遍相信却从未证明过的一些猜想，但总还有弦理论家相信它们。一定的乐观当然是好的，但当结果完全错误时，那就不好了。遗憾的是，图书、文章和电视节目通常展示给公众——也包括科学家——的图像，完全不同于我们直接从发表的结果解读出来的东西。例如，在一本物理学家的流行杂志上，有篇苏斯金2005年《宇宙景观》一书的评论，评论者针对存在多个弦理论的事实说：

这个问题由M理论解决了，那是唯一的包罗万象的理论，通过十一维时空和高维延展的膜包容了那五个弦理论。在M理论的诸多成就中，包括对霍金在20世纪70年代通过宏观论证预言的黑洞熵的第一个微观解释……M理论的问题在于，尽管它的方程是唯一的，却有着亿万个不同的解。[124]

这里最惊人的夸张是说M理论仿佛不是一个建议，而是已经作为精确理论存在了，甚至还有了确定的方程，这没有一句是对的。说什么解释了黑洞熵，也是夸张（第九章已经说过），因为弦理论的结果只适用于特殊的非典型的黑洞。

在专门向公众介绍弦理论的网页上也可以看到这样的歪曲，例如下面的一段：

甚至还有描述引力子（携带引力的粒子）的模式，这也是弦理论受到众多关注的重要原因之一。关键是，我们可以在弦理论中说明两个引力子之间的相互作用，而那在量子场论中是做不到的。没有无穷大！引力不是我们强加的东西，而是在弦理论中自然存在的。所以，弦理论的第一个伟大成就是给出了一个和谐的量子引力理论。125

负责这个特别网页的剑桥弦理论家们知道，没人证明过“没有无穷大”。但他们对猜想很自信，所以把它作为事实呈现给大家。而且，他们也提出了五个不同超弦理论的问题。

这时人们才意识到那五个弦理论其实是同一颗行星上的几个岛屿，而不是不同的东西！于是，存在一个基础的理论，所有弦理论不过是它的不同方面。那就是所谓的M理论。M可以代表所有理论之母，也可以代表神秘，因为我们说的M理论的行星几乎还无人开拓。

虽然最后一句承认M理论“还无人开拓”，但他们还是明确说了“存在一个基础的理论”。普通读者会根据这些话得出结论说，有一个M理论，它具备一般理论的属性，是以精确原理和精确方程建立起来的。126

很多评论文章和谈话对弦理论的结果也说着同样似是而非的话。人们实在弄不清弦理论到底完成了什么，只是一味地夸大结果，缩小困难。我问过一些专家关于那几个重要猜想（如微扰的有限性、S对偶、马尔德希纳猜想、M理论）的现状，惊讶地发现许多弦理论家都不能准确而详细地回答我的问题。

我明白这些责难很重，还是拿一个例子来说明吧。弦理论的一个基本主张是它是有限理论。这意味着它对所有物理学问题的答案都只包含有限的数。显然，任何可靠的理论必须对概率问题做出有限的回

答，对某个粒子或力的强度做出有限的预言。然而，人们提出的关于基本力的量子理论经常不能做到这一点。实际上，在满足相对论原理的众多力的理论中，除了极少数的几个以外，对这些问题都只能给出无限的答案。引力的量子理论尤其如此。许多前景看好的方法就因为不能给出有限的答案而被抛弃。少数的例外包括弦理论和圈量子引力。

正如我在第十二章讨论的，只有在一定的近似方法（叫弦微扰理论）下，才能说弦理论给出有限答案。这项技术产生了无限多个给定环境下的弦运动和相互作用的近似。我们可以讨论一阶近似、二阶近似、17阶近似、1亿阶近似以至无穷阶。为了在这样的系统下证明一个理论是有限的，必须证明它的每一项都是有限的。这很难做，但也不是不可能。例如，电磁学的量子理论（即量子电动力学QED）在20世纪40年代末和50年代初就做好了。这是费曼、戴森和他们那一代人的胜利。粒子物理学的标准模型的有限性也由特胡夫特在1971年证明了。

人们在1984～1985年的巨大兴奋，部分原因也是原来那五个超弦理论被证明在一阶近似下是有限的。几年后，著名理论家曼德尔斯塔姆发表了一篇文章，人们认为它证明了所有的项都是有限的。[127]

当时，对曼德尔斯塔姆的论文众说纷纭。其实有一个直观的论证——弦理论家大概相信——有力地说明，如果理论存在，它就会给出有限的答案。同时，我认识的几个数学家精通这门技术所涉及的问题，否定了那个论证是完整的证明。

我多年没听说有限性问题了。随着领域转向其他问题，它几乎完全消失了。网上时常出现讨论这个问题的论文，但我没留意。实际上，直到最近，我都不记得自己对理论的有限性有过什么怀疑。我20年来紧跟的多数进展和我本人在这个领域的许多工作，都是基于假定弦理论是有限的。这些年，我听过许多弦理论家的讲话，一开始就声

称弦理论给出了一个“有限的量子引力理沦”，然后才具体谈时下感兴趣的问题。许多为公众写的书和发表的谈话都断言弦理论是合理的量子引力理论，还或明或暗地表示理论是有限的。就我自己的研究来说，我曾相信弦理论已经被证明是有限的了（或几乎被证明为有限了，只是还i需要补充一些只有数学家才关心的技术细节），这也是我继续对它感兴趣的一个主要原因。

2002年，我应邀为惠勒的纪念会写一篇文章，评述整个量子引力领域，因为惠勒是它的创始者之一。我认为评述这个学科的最好办法，就是把迄今为止各种方法所得到的所有重要结果罗列出来。我希望做一番客观的比较，看看不同方法在同一个量子引力理论目标的驱动下有着怎样的作为。我写了一个草稿，当然，其中的一个结果就是超弦理论的有限性。

为了完成论文，我当然需要引用一些说明那些结果的论文。多数文章都没问题，但在找泫理论有限性的证明时，我却陷入了困境。找遍文章来源，我只找到了曼德尔斯塔姆的原始论文——就是数学家们告诉我论证不完整的那一篇。我也发现了其他几篇讨沦这个问题的论文，但没有一篇有最后的结论。于是我向熟悉的弦理论家发私人邮件，问他们有限性的现状如何、我从哪儿能找到证明的论文。我问过十几个弦理论家，有老的也有年轻的。回答我的几乎每个人都告诉我结论是正确的。多数都没有证明的文献，而引用的也都是曼德尔斯塔姆的那篇。我很泄气，就去找评论文章——那些专门为考察主要结论而写的文审。我查阅了15篇评论，多数都说（或暗示）理论是有限的，[128]看它们引用的文献，只有以前的评论和曼德尔斯塔姆的论文。我真的找到了一篇评论，是俄罗斯物理学家写的，它解释了结果是未经证明的。[129]但我很难相信他是正确的而那些更有名的人（多数是我认识和钦佩的）写的评论却是错误的。

最后，我问了圆周的同事迈尔斯（Robert Myers）。他以一贯的坦诚告诉我，他不知道有限性是不是已经完全证明了，不过他说有个

叫德霍克的人可能证明过它。我开始去找，终于发现德霍克和蓬（Phong）在2001年就已经成功证明了直到二阶近似的有限性（见第十二章）。从1984年到那时，17年里几乎没有什么重大的进步。（我在第十二章讲过，自德霍克和蓬的文章以来的4年里有过一些进步，主要贡献者是贝科维茨。但他的证明依赖于额外的未经证明的假定，所以，尽管它向前迈出了一步，但还不是有限性的完全证明。）因此，实际情况是，在近似的无穷多项中，只有前三项已知是有限的。除此而外，不论过去还是现在，我们都不知道理沦是否有限。

当我在评论中描述这种状况时，别人都不相信。我收到几封邮件，有的还不太礼貌，都宣称我错了，而理论真是有限的，曼德尔斯塔姆已经证明了。我与弦理论家们谈话，也有相同的经历；听说有限性证明从来就没完成，他们多数都很震惊。但他们的震惊还算不得什么；我曾与弦论之外的物理学家和数学家们谈过，他们更加震惊；他们原来也相信弦理论是有限的，因为他们听说是那样的。对我们所有的人来说，正因为以为弦理论是有限的，才把它看得那么重要。我们从来没听哪个弦理论家指出过它的有限性还是一个尚未解决的问题。

要我写一篇文章来详细评述支持弦理论各种猜想的证据，我也觉得有点儿特别。当然了，我想这应该是学科的领导者们经常做的事情。这种批评，特别是针对未解的问题，在量子引力、宇宙学和其他很多科学领域（我想）都是很寻常的。现在，这件事情不是让弦理论的任何领导者来做，而把责任推给了像我这样的半个内行——虽然明白它的技术，却没有它的社会学义务。我也必须做，因为我对弦理论怀有兴趣，而且那时一直在为它工作。不过，有些弦理论家认为我的评论是一种敌对行为。

马赛理论物理中心的罗维里是我在量子引力的好朋友。他曾把弦理论并未证明有限的结果写进一篇对话，戏剧性地表现不同量子引力方法之间的区别，也遭遇过同样的经历。他收到很多邮件声称曼德尔斯塔姆已经证明了理论是有限的，于足他干脆决定写信问问曼德尔斯

塔姆本人有什么看法。老曼退休了，但很快回了信。他解释说他证明的是某些类型的无穷大项永远不会出现在理论的任何地方。但他告诉我们，他其实并没有证明理论本身是有限的，因为可能出现其他类型的无穷大项。迄今还没见过那样的项，但也没人证明它们不可能出现。

我和许多弦理论家讨论过这些问题，他们在听说理论没有被证明为有限时，也没有一个人决定停止弦理论的研究。我也遇到过一些著名弦理论家，他们坚持在几十年前就已经证明了理论的有限性，只是因为存在某些未解的技术问题才没有发表出来。

但是，如果有限性问题解决了，那我们就要问，一个研究项目有那么多的人，怎么会都不知道他们领域的那个关键问题的现状呢？在1984年和2001年间，许多弦理论家在讲话和文章里都将弦理论是有限的作为一个事实，这难道不应该关注吗？为什么众多弦理论家都心安理得地向圈内外的人宣扬，以模糊的语言暗示理论是完全有限而和谐的？

下面是曼德尔斯塔姆2006年6月8日的电子邮件：

关于我对n圈弦振幅的有限性的论文，我首先要说的是，发散只能出现在模空间退化的地方。我考察过伴随“膨胀子”发散的退化点，那是与弦理论有关的。我证明了以前用于单圈振幅的论证也可以推广到n圈振幅，而且偶超模上的积分围道的定义问题也可以通过与幺正性相容的唯一法则来解决。我同意这并未严格地在数学上证明有限性，但我相信它处理了可能引起无穷大的物理学问题。我没有考察从以前的对偶模型了解的其他无穷大的来源，即虚时间的运用。如果在虚时间积分，因子exp（iEt）（其中E是邻近能量与初始能量之差）显然是可能发散的。人们根据物理背景相信，这种无穷大可以通过向实时间的解析延拓而去除，这在无圈和单圈振幅情形已经有了具体证明，而且还证明，得到有限结果的解析延拓在两圈振幅的情形也能定

义。

（在本书的电子本中，作者引用的是曼德尔斯塔姆2003年11月5日的回答：你问我是否同意多圈振幅的有限性猜想仍然悬而未决。如果你说“未决”是指没有相反情形的严格证明，那我没意见。我在前一封信里说过，我并不说整个处理都是严格的，对开弦理论而言，可以说有部分工作还没做，连不严格都谈不上然而，我想多数物理学家会认为“未决”并不仅仅指相反的证明不严格；否则物理学就该有太多的“未决”问题了！“猜想”一词似乎意味着多圈振幅的问题没有真正经过检验，只是基于单圈图才相信它们是有限的。事情当然不是那样的。我们知道发散可以出现在模空间退化的地方（这是严格的）。我们也知道模空间如何退化，而且考察过退化区域。当然也可以指出考察中有些地方不够严格。我已就自己的理解谈了这种状况；大概不能将其简单归结为“有限性在所有阶都证明了”或“只能猜想单圈以上的有限性”。——译者）

除了有限性问题，弦理论还有一些普遍相信但迄今尚未证明的猜想。我们已经讨论过，马尔德希纳猜想在文献中有几个不同的形式，它们的意义差别很大。我们确定的是，猜想的最强形式远未证明，尽管有些弱形式得到了很好的支持。但弦理论家并不这么看。在最近关于马尔德希纳猜想的一个评论中，霍洛维茨和波尔琴斯基将它比作数学中著名的猜想——黎曼猜想。

总之，我们有很好的理由将[马尔德希纳对偶猜想]列入正确但未证明的范畴。实际上，我们认为它与黎曼猜想那样的数学猜想属于同样的东西。两者都在看似不同的结构间建立了联系……而且，尽管每个都是注意的焦点，但仍然既没证明也没否定。130

我从没听数学家说一个结果是“正确但未证明”的，除此而外，这句话的惊奇在于两个聪明的作者忽略了那两个猜想的一点显著区别。我们知道，黎曼猜想联系的两个结构（素数和某些函数）都是在数学

中存在的东西；问题只是它们之间的那个猜想关系。但我们不知道弦理论或超对称规范理论是否真的以数学结构的形式存在；实际上，它们的存在也是问题的一部分。这段话清楚地说明，作者在推论时假定了弦理论是一个确立好了的数学结构——似他们忘了，即使它是真的，我们也一点儿不知道那结构是什么。如果不做这样的未经证明的假定，那么你对最强形式的马尔德希纳猜想的证据的评价就不会和他们的一样。

弦理论家为他们相信这些未经证明的猜想辩解时，常常会说那是弦理论群体“普遍相信的”，“没有哪个理性的人会怀疑它是真的”。他们似乎觉得求助他们圈内的共识就等于合理的推论。下面是一个典型的例子，来自一个著名弦理论家的博客（引文中的黑体字是我强调的）：

在过去六年里没有睡大觉的人都知道在渐近反德西特空间里的量子引力具有幺正时间演化……面对那么多AdS/CFT的证据，我想那些不相信上述论断不仅在霍金考虑的半经典极限下成立而且在整个非微扰理论也成立的人，恐怕没有几个还会坚持己见了。[131]

被迫承认自己属于顽固分子，感觉并不舒服，但经过对证据的详细考察，我不得不承认这一点。

漫不经心地对待关键猜想的证据，在几个方面起着阻碍作用。首先，连同前面说的那些倾向，它意味着几乎没人会研究这些重要的开放问题——这就可能使它一直悬而不决。它还会腐蚀科学的道德和方法，因为一大群聪明人宁愿相信猜想也不想去寻找它们的证明。

而且，在发现重大结果时，他们常常夸大它的意义。几个非弦理论家曾问我，既然弦理论已经彻底解释了黑洞熵，为什么不去做别的呢。虽然我非常佩服斯特罗明戈和瓦法等人关于极端黑洞的研究（见第九章），但我必须重申，似乎有很好的理由证明，精确结果不能推

广到一般的黑洞。

同样，关于大量弦理论（即众说纷纭的“理论景观”）都有正宇宙学常数的论断也远未确定。不过，某些一流弦理论家还是凭着这些软弱的结果大肆宣扬弦理论的成功和美好的前景。

一贯的夸张也许给弦理论带来了比它的对手更多的好处。弦理论家总是声称他们的研究要解决领域的重大问题，而别的科学家却只能说有证据证明存在某个理论——迄今尚未建立的理论——可能解决那些问题。如果你是大学的系领导或出资单位的官员，岂不也是更愿意资助那些要解决大问题的项目、聘用为那些项目工作的人吗？

让我来总结一下，看看它将我们引向何处。以上的讨论暴露了弦理论群体的7个异乎寻常的特征。

1.极大的自信，从而自以为高人一等，是精英里的精英。

2.异常统一的群体，不论证据强弱，都有强烈的舆论意识，对开放问题有异乎寻常的一致的观点。这些观点似乎关联着一种等级结构，几个领导者的思想指引着领域的观点、策略和方向。

3.群体意识，在某种意义上类似于宗教信仰或政党纲领的认同。

4.强烈的界线意识，将群体与其他专家分隔。

5.漠视本群体外的专家的思想、意见和工作，只愿在群体内部交流。

6.乐观的倾向，过分解释证据，相信夸大或错误的结果，拒不考虑理论可能是错误的。这应和着另一种倾向：相信某个结果是因为“大家都相信它”，即使没人检验过（甚至见过）证明本身。

7.对研究计划应该考虑的风险程度缺乏认识。

当然，并不是所有弦理论家都这样，但在弦理论群体内外，几乎没有人会否认这些态度刻画了多数弦理论家的特征。

我要明白地说，我不是在批评个别人的行为。许多弦理论家个人都很谦虚，能自我批评。如果问起来，他们会说他们也为群体的这些特征感到难过。

我还要说明，我也和弦理论的同事一样老犯错误。多年来，我一直相信有限性那样的基本猜想是证明了的。这是我用多年时间做弦理论的主要原因。除了我自己的工作受影响而外，在量子引力群体中，我也是最强力地为弦理论摇旗呐喊的人。我还没时间检阅文献，所以，我也乐意请弦理论群体的领导者们来批判我的思想。在我做弦理论的那些年里，我非常在乎群体的领导者们对我的工作的看法。我也和任何青年一样，盼望我那小圈子里的大人物能接受我。如果说我没有真的听取他们的忠告，全身心投入那个理论，那只是因为我个性太倔，在这种情况下老是转不过弯来。对我来说，这不是“我们”与“他们”的问题，也不是两个群体之间为了争先的决斗。这些都是我个人的问题，自从我做科学以来就一直在与它斗争。

所以我非常同情弦理论家的尴尬境遇，他们既想做好科学家，又想证明自己在本领域的影响力。我很理解，要群体接受你，你就必须相信一大堆自己都不知道如何证明的思想，那时要清楚而独立地思考，当然是很困难的。这是我想了多年才跳出的陷阱，所有这些都令我相信我们物理学家走入困境了。如果你问弦理论家，为什么从来不请研究其他可能方法的人参加弦理论会议，他们很多人会赞同你的意见，说应该请那些人。他们也为眼下的状况感到难过，但会坚持说在这方面无能为力。如果你问他们为什么弦理论群体从来不请研究其他方法的年轻人做博士后、教员或访问学者，他们也会赞同你，说那是好事情，他们很遗憾现在还没有那么做。这种状况存在着大问题，每

个人都认错，但没有人来负责。

我很信任我的弦理论朋友们。我相信他们每个人都很谦虚，都有自我批评精神，都不像他们的群体那样跋扈。

既然每个人都怀着良好的愿望和正确的判断力，群体行为怎么会那样荒唐呢？

原来，社会科学家早就认识了这种现象。它折磨着一些名人的群体，他们只是偶尔看条件才和圈内的人交流。情报和决策机构以及一些大公司研究过这种现象。因为后果有时可能是悲剧性的，有大量文献描述这种现象，称之为小团体思维。

耶鲁心理学家詹尼斯（Irving Janis）在20世纪70年代提出了小团体思维的名词，将它定义为一致思维模式，“当人们置身一个团结紧密的小团体，为了维护团体的一致而不能现实地评估不同的行为过程，就陷入那样的思维模式。”[132]根据这个定义，小团体思维只出现在有着高凝聚力的群体。它需要群体成员有强烈的“团结如一人”的感觉，而且愿意不顾一切代价维护这种群体关系。当同事们都以小团体思维模式行动时，他们会自然将“维护群体和谐”的标准用于所面对的每个决定。[133]

詹尼斯研究过专家群体决策失败的案例，如猪湾事件。从此，“小团体思维”一词就用于许多其他例子，包括NASA阻止“挑战者”灾难的失败，西方预测苏联解体的失败，美国汽车公司对小汽车需求预测的失败，以及最近——也许最灾难性——的失败：布什政府基于伊拉克有大规模杀伤性武器的错误信息贸然发动战争。

下面是俄勒冈州立大学公共网站对“小团体思维”的描述：

小团体思维的成员将自身看作群体的一部分，抵抗反对他们目标的外来者。患有“小团体思维症”的人都有下面的特征：

1.过高估计群体的抵抗力和道德水平。

2.集体合理化决策。

3.丑化和僵化外来群体及其领导者。

4.一致性的群体文化，个人之间相互监督，以维护群体的完全一致性。

5.某些成员想方设法向领导者隐瞒自己和其他群体的信息，以维护群体领导的权威。[134]

这和我刻画的弦理论文化没有一一对应，但二者的相似足以令人忧虑。

当然，弦理论家很容易回应这种批评。他们可以举出很多历史的事例，说明科学的进步依赖于专家群体形成团结一致的意见，而外来者缺乏足够的评估证据和进行判断的职业技能，所以他们的意见必须抛弃。由此，科学群体必须拥有树立和加强共识的机制。在外行人看来那也许像小团体思维，实际上是理性的，是在严格的约束规则下运行的。

他们还可以反驳人们谴责他们以群体的意见取代了个人的重要思想。我和一个著名的科学社会学家讨论过这个问题，他说未经证明就相信一些关键猜想并不稀奇。[135]没有哪个科学家能直接证明构成其学科信仰基础的所有实验结果、计算和证明；几乎没人有那个能力，而在当代科学中，谁也没有那么多时间。于是，当你加入某个科学群体，就必须相信同事告诉你的他们专业领域的结果。这就可能导致将猜想作为事实来接受，但它既发生在最终成功的研究中，也同样多地发生在失败的项目里。当今科学如果没有一群可以相互信任的人，简直就无法进行下去。因此，虽然这样的插曲令人遗憾，而且在出现时就该修正，但它们本身并不意味着什么注定失败的研究或病态社会

学。

最后，老牌的弦理论家可以说他们德高望重，有资格在他们觉得合适的地方指导研究。毕竟，科学实践以直觉为基础，而这就是他们的直觉。会有人浪费时问做他们不相信的事情吗？他们当然只会请人来做那些他们认为最可能成功的理论。

那么如何同答这种辩解呢？如果科学基于专家群体的共识，那么你在弦理论中所拥有的就是一个群体，其中的专家对他们研究的理论的正确有着惊人一致的意见。有什么合理的根据——不论以什么方式——拿出理智而切实的反对意见吗？我们需要做很多事情，而不仅仅是抛弃“小团体思维”这样的字眼。我们必须有一个关于科学是什么和如何运行的理论，它将清楚地说明，如果一个特殊的群体在理沦未经正常检验之前就在领域中占据主导地位，那是科学的悲哀。这是我们现在面临的使命。

# 第十七章　什么是科学

为了扭转物理学的这股恼人的潮流，我们必须首先明白什么是科学——什么推动它向前，什么拉着它后退。为此，我们对科学的定义必须超越所谓科学家所做的事情的总和。本章的主要目的就是提出这样一个定义。

我1976年走进哈佛大学研究生院时，还是来自小学校的天真学生。我敬畏爱因斯坦、玻尔、海森伯和薛定谔，惊讶他们神奇的思想力量给物理学带来的变革。和许多年轻人一样，我做梦都想成为他们那样的人。这时我置身于粒子物理学的中心，周围都是领域里的头面人物——如科尔曼、格拉肖和温伯格。这些人聪明绝顶，但一点儿也不像我心目中的英雄。上课时，我从没听他们讲过空间和时间的本质或量子力学的基本问题。我也没见过有多少学生对这些问题感兴趣。

这使我陷入了危机。我当然不如来自名校的同学那么基础扎实，但我在读大学时就已经做过研究，而多数同学都没做过；我也知道我学得很快，所以我自信能做物理学的工作。但我对如何才算一个伟大的理论物理学家也有特别的想法。我在哈佛遇到的那些大理论物理学家和我的想象相距甚远。那里的氛围严酷而好斗，没有一点儿哲学味道，尽是些冲动、高傲、自负的人，还时常伤害与他们意见相左的人。

这期间我和年轻的科学哲学家雷切尔一科恩（Amelia Rechel-Cohn）成了朋友，通过她结识了和我一样对哲学和物理学基本问题感兴趣的人。但这使事情变得更糟。他们比理论物理学家好一点儿，但似乎只乐于分析狭义相对论或普通量子物理学基础的逻辑问题。我对那样的谈话毫无耐性；我想创造理论，而不是批评理论。我确信——标准模型的创立者们似乎也曾那么草率——他们知道我需要知道的事情。

正当我认真考虑放弃时，阿米丽亚（Amellia）给了我一本哲学家费耶阿本德（Paul Feyerabend）的书。书名叫《反对方法》，就像在对我说话——但它说的不是很令人鼓舞。它对我的天真和专注是一个打击。136

费耶阿本德在书中对我讲的是，看哪，孩子，别做梦了！科学不是坐在云端里的哲学家。它是人的活动，与任何别的东西一样复杂，一样成问题。科学没有单一的方法，谁是好科学家也没有单一的标准。好科学就是在历史的特定时刻增进我们知识的东西。别来烦我如何定义进步——随便你用什么方式定义，都是对的。

从费耶阿本德那儿我认识了进步有时需要深刻的哲学思想，但多数时候都不是那样的。进步的实现多数是靠投机者们抄捷径、夸大他们的知识和成就。伽利略算其中的一个；他的许多论据都是错误的，而他的对头——当时受过良好教育、善于哲学思维的耶稣会士天文学

家们——很容易发现他的思想漏洞。不过，最终是他对了而他们错了。

我从费耶阿本德认识了没有什么先验的论证能告诉我们什么东西能适应所有的环境。某个时刻推进科学的力量在其他时候可能就是错的。从他讲的伽利略的故事，我还明白了更多的东西：你必须为自己的信仰而斗争。

费耶阿本德的言论远非及时的清醒剂。如果我想做好科学，我就必须认识到我有幸合作的人都是当代的大科学家。他们和所有大科学家一样，是靠正确的思想和奋斗取得的。如果你的思想正确并且为之奋，总会取得成绩的。不要浪费时间替自己难过，也别为爱因斯坦和玻尔伤感。没人能帮助你，只有自己能发展自己的思想，也只有自己能为它们奋斗。

我走过很长的路才决定留下做科学。我很快发现，将粒子物理学用过的方法用于量子引力问题，是做不了真正研究的。如果这意味着暂时将基本问题放在一边，那么能建立新的基础并在新基础进行计算，也是了不起的事情。

为了感谢费耶阿本德挽救了我的事业，我给他寄了一本我的博上论文。他回信时给我寄了一本他的新书《自由社会的科学》（1979），还请我去伯克利时访问他。几个月后，我正好去加州参加粒子物理学会议，就设法去找他，可是去得太突然了。他不在学校办公，连办公室也没有。当我打听他时，哲学系秘书笑了，让我去他家找。电话本上有他的地址，在伯克利山米勒大街。我鼓起勇气拨通了他的电话，礼貌地说要找费耶阿本德教授。不知谁在电话那头大声说，“费耶阿本德教授！那是另一个人。你可以在学校找他。”然后就挂了。于是我到他的班上去找到他了，然后进行了友好的谈话，可惜时间太短。可就在这几分钟里，他给了我一个无价的忠告。“是的，学术界一团糟，你做不了什么事情。可是别担心。就做你想做的。如

果你知道你想做什么，并且大力倡导，没人能阻止你做下去。”

半年后，他给我写了第二封信，寄到圣塔巴巴拉，我刚去那儿的理论物理研究所做博士后。他说他和一个有才干的物理学本科生谈过话，那人和我一样也对哲学感兴趣。问我是否愿意见他，给他提一些建议。我想的是能有第二次机会和费耶阿本德谈话，就到了伯克利，在哲学楼的阶梯上见到了他们两位（显然就像他和同事一样亲近）。费耶阿本德请我在“加州料理”（Chez　Panisse）吃午饭，然后带我们去他家（原来就是在伯克利山米勒大街），这样他就能看他喜欢的肥皂剧，而那个同学和我也能谈话。在路上，我和费耶阿本德坐在他的小跑车的后排。他的车装了充气筏，即使发生八级地震，他也能安全通过海湾大桥。

费耶阿本德提出的第一个问题是重正化，那是量子场论中处理无穷大的一种方法。我惊奇地发现他非常熟悉当代物理学。他并不像我的某些哈佛教授说的那样，他不反科学。很明显他喜欢物理学，他比我见过的多数哲学家都更喜欢谈技术性问题。他作为科学敌人的名声无疑是因为他考虑了为什么科学会没有结果的问题。那难道是因为科学有方法吗？巫医也有方法啊。

我冒昧地认为，也许区别在于科学运用了数学。他回答说，占星学也用数学，而且，如果需要的话，他可以解释占星学家们所用的不同计算体系的细节。他举例说，开普勒（历史上最伟大的物理学家之一）为占星学的技术进步做出了几个重大贡献，而牛顿在炼金术上花费的时间比物理学还多，对这些事情我们都不知道该说什么。难道说我们是比开普勒和牛顿还伟大的科学家吗？

费耶阿本德相信科学是一种人类活动，是投机者的事业，他们不遵从一般的逻辑或方法，而做任何能增进知识（不论你怎样定义）的事情。所以，他的大问题是：科学如何运作，它为什么运作得那么好？即使他反对我的所有解释，我觉得他热情追求这个问题，不是因

为他反科学，而是因为他关心科学。

然后，费耶阿本德给我讲了他的故事。他十几岁在维也纳时很有物理学天赋，但他应征参加第二次世界大战时，研究活动也缩减了。他在俄国前线受伤，后来在柏林退伍，战后在那儿谋得一个演员的差事。不久，他对剧场厌倦了，又回到维也纳做物理学研究。他参加了哲学俱乐部，发现只需要简单发挥从表演专业学来的技巧，在争论的任何一边都能赢。这使他怀疑学术的成功是不是还有任何理性基础。一天，学生们把维特根斯坦（Ludwig Wittgenstein）请来俱乐部。费耶阿本德被深深触动了，决定走进哲学。维特根斯坦和他谈话，请他去剑桥一起做研究。但当他到英国时，维特根斯坦去世了，于是有人建议他和另一个从维也纳流亡出来的波普尔（Karl Popper）谈谈，他正在伦敦经济学院教书。因此他到了伦敦，从写攻击波普尔著作的论文开始了他的哲学生涯。

几年后，他得到一个教师职位。他问一个朋友，自己知道的很少，应该怎样上课。朋友告诉他，把他认为知道的东西写出来。他写满了一页纸。朋友接着告诉他，拿第一句话做第一堂课的主题，第二句话做第二课的主题，依此类推。就这样，这位从物理学学生经过士兵和演员的人，成了哲学教授。137

费耶阿本德开车送我们回伯克利校园。离开时，他给我们提出了最后一个忠告。“就做你想做的，不要管别的事情。在我的经历中，从来没有花过五分钟做我不想做的事情。”

我多少也是这么做的。一直到今天。现在我觉得我们不但要谈科学思想，还要谈科学过程。这是别无选择的。我们的后辈会接着思考为什么我们远不如我们的老师那么成功，我们对他们是负有责任的。

自从我访问费耶阿本德（他在1994年去世，70岁）以来，我引导了几个天才的青年学生走出我经历过的那种危机。但我不能对他们

说我当年对自己说过的话——那时主流的研究风格成就辉煌，是必须尊重和服从的。现在我不得不同意我的青年同事，主流的风格是不成功的。

首先，也是最重要的，我在哈佛大学学习的那种做科学的风格没有让我取得进步。它建立标准模型成功了，却没能超越。30年过去了，我们必须问它如今是否已经落伍了。也许眼下需要爱因斯坦和他的朋友们的那种更思辩、更冒险也更哲学的风格。

这个问题远比弦理论广阔；它涉及整个物理学共同体多年形成的价值和态度。简单说，物理学共同体是以大项目的飞扬跋扈和小项目的谨小慎微构建起来的。于是，年轻的科学家如果能为大项目的问题提出技术精妙的解，令老科学家感到满意，他就有机会成功。如果反其道而行——独立思考并试图构建自己的思想——就只能是悲剧。

于是，物理学不能解决本身的关键问题。现在是转折的时候了——鼓励小的、冒险的、新的研究计划，抵制顽固的老方法。我们应该让爱因斯坦们重新回到主流，思考自己的问题，不要管强力科学家们的既有思想。

但为了说服怀疑者，我们必须回答费耶阿本德关于科学如何运作的问题。

似乎存在两种对立的科学观。一种认为科学是叛逆者的领地，所谓叛逆者，就是那些有着宏大新思想的人，他们奋斗一生就为了证明自己是正确的。这是伽利略神话，今天在几个令人敬慕的大科学家的奋斗中，如数学物理学家彭罗斯、复杂性理论家考夫曼（Stuart Kauffman）和生物学家马格里斯（Lynn　Margulis），还能看到那神话的影子。另一种观点则把科学看成保守的、一致的共同体，几乎不能容忍丝毫的对正统思想的背离，而将创造的能量输入业已确定的研究纲领。

从某种意义说，两种观点都是对的。科学既需要叛逆者也需要保守者。乍看起来者似乎矛盾。一个蓬勃发展多个世纪的事业怎么会让叛逆者与保守者共存呢？诀窍是这样的：让叛逆者与保守者一辈子在科学共同体里做对头，在一定程度上也是在个人的内心里做对头，一辈子不得安宁。但这是如何做到的呢？

从每个科学家都有自己的声音看，科学是民主，但它绝没有少数服从多数的法则。而且，虽然每个人的判断都得到了重视，但共识起着决定性的作用。当我的专业的大多数人都去迎接一个我不能接受——即使有好处也不愿接受——的研究项目，我还能有什么落脚的地方吗？答案是，民主不仅是多数人的法则。在多数人法则之外还有着意识和伦理的系统。

于是，如果我们要说科学不仅是社会学，不仅是学术政治，那么我们必须要有一个关于科学是什么的概念——要符合（而且超越）科学是人类的一个自治群体的思想。要说哪个特殊的组织形式或特殊的行为对科学好或坏，我们必须有一个超越普遍事物的价值判断的基础。我们必须有一个能脱离多数而不会被说成是怪人的基础。

我们先把费耶阿本德的问题打碎成几个简单的小问题。我们可以说当科学家对问题达成共识时科学就进步了。这种事情发生的机制是什么呢？在共识达成之前常有争论。不同的意见对科学进步起着什么铺路的作用呢？

为回答这些问题，我们应该回到早年的哲学家们的观点。20世纪20年代和30年代，维也纳发生了一场叫逻辑实证主义的哲学运动。逻辑实证主义者提出，断言只有在被世界的观测所证实后才成为知识。他们称科学知识是这些被证实的命题的总和。当科学家做出能检验而且确实得到检验的断言时，科学就进步了。他们的动机是摆脱形而上学的哲学，那种哲学已经充斥了空洞无物的大书。在这一点上他们部分成功了，但他们谨慎的科学特征没能持续多久。很多问题出现了，

其中一个就是，在观察与陈述之间不存在铁定的对应。最简单的观察里也溜进了假定和偏见。将科学家的言论打碎成与被剥夺了理论的观察相对应的小原子，是不现实的，甚至也许是不可能的。

当实证论失败时，哲学家提出科学进步是因为科学家走上了通向真理的路线。科学方法的建议是卡尔纳普（Rudolf Carnap）和奥本海默（Paul Oppenheim）等人提出来的。波普尔也提出了自己的建议，那就是科学进步在于科学家提出了可以证伪的理论——就是说，他们做出了可以被实验否定的陈述。在波普尔看来，理论永远不可能被证明是正确的，但如果它经受了很多试图否定它的试验，那么，我们就开始有了它所包含的真理——至少在它最后被否定之前。[138]

费耶阿本德的哲学起步就是攻击这些思想。例如，他证明否定一个理论并不是那么容易的事情。很多时候，就在一个理论似乎已经被否定以后，科学家还会坚守它；他们只需要改变实验的解释就能做到这一点。他们也会挑战自己的结果。有时这会终结一个理论，因为它确实错了。面对明显的实验矛盾而坚守一个理论，有时竟然是正确的做法。你怎么知道自己处于哪种状况呢？费耶阿本德认为，你不可能知道。不同的科学家抱着不同的观点，让运气来证明自己。没有一个普遍的法则能告诉我们应该在什么时候抛弃或维护一个理论。

费耶阿本德证明，科学家会在紧要关头打破法则而取得进步，这就从总体上攻击了方法决定科学进步的思想。而且，他还指出——在我看来是令人信服的——如果总是遵照“方法的”法则，科学可能停滞不前。科学史家库恩从另一个方面攻击了“科学方法”的思想，他说科学家在不同的时候用着不同的方法。但他不如费耶阿本德激进，他想提倡两个方法：“正常科学”的方法和科学革命的方法。[139]

另一个批判波普尔思想的是哲学家拉卡托斯，他指出在证伪与证实之间不存在像波普尔假定的那么大的不对称。如果你看见一只鲜红的天鹅，你也不大可能放弃说所有天鹅都是白色的理论；你反而会去

找出一个为它涂颜色的人。[140]

这些论证给我们留下几个问题。第一是科学的成功还需要解释；第二是（波普尔强调的）不可能区分物理学和生物学等科学与其他自称为科学的信仰体系——诸如马克思主义、巫术和智能设计。[141]如果没有这种区分，那么可怕的相对主义就可能溜进来，一切关于真理和实在的断言都一样有道理。

虽然我和许多实践的科学家一样相信我们不追随单一的方法，我也相信我们必须回答费耶阿本德的问题。我们可以从科学在人类文化中的作用说起。

科学是人类文化的工具之一，自史前时代起，它就随人类自身所处状况应运而生：我们可以梦游无限的空间和时间，无限的神奇和无限的美好，终于发现我们身处几个不同的世界：物理世界、社会世界、想象世界和精神世界。我们之所以成为人，是因为我们在长期寻找工具，让它赋予我们把握形形色色世界的力量。这些工具现在叫科学、政治、艺术和宗教。现在，它们像远古时代一样，也赋予我们把握生命的力量，形成我们希望的基础。

不论它们叫什么，从来没有哪个人类社会没有科学、政治、艺术和宗教。在石壁画着古代猎人的洞穴里，我们发现了有一定模式的石头和骨头，说明古人以14、28或29为一组计数东西。考古学家马沙克（Alexander Marshack，《文明之根》的作者）将这些解释为月相观测。[142]它们也可能是早期节育方法的记录。不论哪种情形，都说明人类在2万年前就用数学来组织和概化他们的自然经验了。

科学不是发明的。随着人们发现了工具，学会了将物理世界带入我们认识领域的活动，科学也随时间进化着。于是，科学之所以成为科学，是因为自然就是那样的——也因为我们就是那样的。许多哲学家错误地去寻求科学为什么卓有成效的解释，还想将它用于任何可能

的世界。但不存在那种东西。如果幻想有什么方法能在任何可能的宇宙发挥作用，那么它就像一把试图让任何动物都舒服的椅子：在多数情形都只能是很不舒服的。

实际上，有可能证明这句话。假定科学家像一群在乡下找最高峰的瞎子。他们看不见，但可以凭感觉确定哪条路向上，哪条路向下。而且他们还有带声音的高度计，可以告诉他们所处的高度。他们在最高处看不见，但他们知道，因为只有在那个地方，所有方向才都会向下。问题是，可以存在多个高峰，如果你看不见，那就难以确定你是不是爬上了最高的那个峰。这样，是否有办法让瞎子在最短的时间里找到最高峰，并不是显而易见的。这是数学家经常研究的问题，最后证明是不可能的。计算机科学家沃尔波特（David Wolpert）和马克雷迪（William Macready）确立的“没有免费午餐”定理说，在每个可能的景观中，没有什么办法能比随机游走的结果更好。[143]想让策略的效果更好，必须知道景观的一些东西。这种策略在尼泊尔也许成功，但在荷兰可能失败。

因此，哲学家不能发现解释科学行为的普遍策略也就不足为奇了。他们确实发明了一些策略，但与科学家的实际作为没有多少相似的地方。成功的策略是随时发现的，它们是与个别的科学实践密切相连的。

一旦我们明白了这一点，我们就能认定科学开拓的那些自然特征。最重要的特征是，自然是相对稳定的。在物理学和化学中，很容易设计一些结果可以重复的实验。情况不一定都这样，例如，在生物学中就不是，而心理学就更不是了。但在实验可以重复的领域，通常都用定律来描述自然。于是，物理学的实践者们从一开始就对发现一般定律感兴趣。这里的问题不在于是否真的存在基本定律；对我们的科学实践来说，关键在于是否存在那样的法则，我们可以靠自己制造和掌握的工具来发现和模拟它们。

我们碰巧生活在一个方便我们理解的世界，而且一直是这样的。从我们以一个物种出现以来，就轻易发现了许多规律，如星空和季节，动物的迁徙与植物的生长，以及人类自己的生物周期。我们学会了通过在兽骨和石头上留下记号来记录这些规律，将它们联系起来，发挥它们的作用。今天，我们用巨型望远镜、强大的显微镜和越来越大的加速器做实验，也不过是在重复我们一直做着的事情：用掌握的技术去发现我们面前的模式。

但如果说科学存在是因为我们生活在一个充满规律的世界，那么它的特殊表现形式正是因为我们自身的特殊性。特别是，我们善于从不完备的信息得到结论。我们不断地观察世界，预言它、判断它。那是猎人的做法，也是粒子物理学家和微生物学家的做法。我们永远也不会有足够的信息来完全地证实我们的结论。要成为一个好的生意人、好的猎人、好的农民和好的科学家，基本的素质之一就是会凭直觉进行猜想，能在有限的信息未形成完整证明的时候自信地行动。正因为这一点，人类才成为如此成功的物种。

但这种本领付出了很高的代价，那就是我们很容易欺骗自己。当然，我们知道很容易被别人欺骗。人们很容易相信谎言，因为它很有效。毕竟我们只能根据不完备的信息得到结论，因而在谎言面前实在软弱无力。我们的基本态度只能是信任，如果要求证明所有的东西，我们就没有什么可以相信的，也就不可能做任何事情——不能走下床，不能结婚、交朋友或加入任何组织。没有信任，我们就会成为孤独的动物。语言之所以行之有效，因为我们多数时候都相信别人告诉我们的东西。

但同样重要而且令人清醒的问题是我们在多大程度上欺骗自己。我们不仅欺骗个人，也欺骗整个群体。群体容易轻信某些在个人后来看来明显是错误的东西，这种倾向确实很可笑。20世纪的一些最悲惨的事情就是因为好人上了坏领导者的当。但达成共识是我们赖以生存的基础，只有那样猎人才能收获，部族才能躲避危险。

所以，一个群体要生存，就必须有纠错的机制：长者让年轻人不冲动，因为他们从自己的漫长经历明白了犯错误是多么容易。年轻人挑战几代人的信仰，因为那些信仰已经落伍了。人类社会能进步，就是因为它学会了既要叛逆也要服从，发现了在漫长岁月里平衡两种品性的社会机制。

我相信科学就是那样的一种机制。它是孕育和激励新知识发现的方式，但在历史的长河里，它更证明是有效的揭露谬误的一整套技艺和实践。为了不断克服我们自欺也欺人的固有倾向，科学是最好的工具。

从这段概说我们可以看到，科学与民主有相同的地方。不论科学群体还是更大的群体，都需要达成结论并根据不完整的信息做出决定。在两种情形，信息的不完整都将引发不同观点的派别。不论科学的还是非科学的社会，都需要解决纷争与调和不同意见的机制。这种机制需要我们揭露错误，允许用新的方法去解决难解的旧问题。人类社会有很多这样的机制，其中有的依靠武力和强权。民主的最基本理念是社会只有在和平解决纷争的情况下才能运行最好。因此，科学和民主都痛感我们欺骗自己的倾向，也都乐观地相信我们作为一个社会能改正错误，使我们在整体上越来越比任何个人更聪明。

现在我们已经把科学置于恰当的背景下了，下面接着讨论它为什么那样有作为。我相信答案很简单：科学之所以成功是因为科学家组成了一个由共同的道德规范树立和维持起来的群体。我相信，正是对一定的道德规范而不是任何特殊事实或理论的忠诚，在科学群体的自我改正中起着根本的作用。

这套规范有两个原则：

1.如果一个问题可以由忠实的人通过将合理论证用于公开证据而确定，则可以认为它的结果必然就是那样的。

2.反之，如果根据公开证据的合理论证不能使忠实的人就某问题达成一致意见，则这将必然允许甚至鼓励人们提出各种不同的结论。

我相信科学成功是因为科学家遵从了（也许不是那么彻底）这两项原则。为说明这一点，我们看看这些原则要求我们做些什么。

●我们赞同根据共享的证据踏实地进行合理论证，而不管结果的一致性程度有多高。

●每一个科学家都自由根据证据得出自己的结论。但每个科学家也需要向整个科学共同体提出这些结论的论证。这些论证必须合理而且建立在所有成员都能获得的证据上。证据、获得证据的方法以及根据这些证据导出结论所利用的论证逻辑，必须面向所有成员的检验。

●科学家根据共享证据演绎可靠结论的能力基于对工具和多年进程的把握。之所以学习那些工具和进程，是因为经验已经证明，它们常引出可靠的结果。每个经过这些方法训练的科学家对错误和自欺的流毒都深有感触。

●同时，科学共同体的每个成员都意识到最终目标是达成共识。共识也许突然出现，也许需要一定时间。科学工作的最终裁判是共同体的未来成员，他们在遥远的未来能更好、更客观地对证据进行评价。虽然科学项目可以暂时团结一些拥护者，但没有一个项目、声明或观点能长久地成功，除非它能产生足够的令怀疑者信服的证据。

●科学共同体平等地面向所有的人。地位、年龄、性别或任何其他个人特征不能影响科学家对证据和论证的考虑，也不能限制他们考虑其他的证据、论证和信息。然而，进入科学共同体需要满足两个条件。首先是至少把握一个科学分支领域的方法并且能独立做出其他成员高度认可的工作。其次是忠实地服从共同的道德规范。

●虽然在某些分支领域会暂时形成正统，但科学共同体认识到，

相互对立的意见和研究项目对共同体的健康发展仍然是必要的。

做科学的人普遍怀有幼稚的欲望，总以为自己正确，总相信自己拥有了绝对真理。但如果加入科学共同体，这些东西就都要抛到脑后。相反，他们应该知道自己是一个正在发展的事业的一员，最终将达到个人不可能达到的目标。他们还要接受专业的训练，在多数情形下学会很多不可能靠个人学会的东西。接着，因为他们为专业实践付出的劳动，共同体保证每个成员有权利宣扬他们认为得到了那些实践的证据支持的任何观点或研究项目。

我愿将这种群体称为伦理群体，其组织取决于对一种行为规范和实现那种规范的专业实践的忠诚。科学在我看来是我们见过的这种群体的最纯粹的例子。

但这还不足以将科学刻画为一个伦理群体，因为有些伦理群体的存在是为了维护旧的知识而不是为了发现新的真理。宗教群体在许多情形都满足伦理群体的标准。实际上，科学的现代形式是从僧侣和神学学院演化而来的——那也是一些伦理群体，其目标是维护宗教教义。所以，如果说我们的科学特征是要树立权威，那么我们必须外加一些能明确将物理学从修道院区分出来的准则。

为此，我想引入第二个概念，即我所谓的想象群体。这种群体的规范和组织体现在它相信进步是必然的而未来是开放的。开放为新生事物和惊奇留下了想象或实在的空间。我们不但相信未来会更好，而且也认识到我们不能预知如何达到那个更好的未来。

不论马克思主义国家还是正统的宗教国家，都不属于想象群体。他们也向往美好的未来，但他们相信他们完全知道怎样达到未来。我小时候常从马克思主义者的祖母和她的朋友们那儿听说，他们确信自己是正确的，因为他们的“科学”教他们学会了“形势的正确分析”。

想象的群体相信未来会带来惊奇，带来新的发现，解决新的危机。他们并不忠实于眼下的知识，而是将希望寄托在子孙后代，将思维的法则和工具传给他们，使他们能利用和超越我们今天无法想象的环境。

优秀的科学家希望学生能超过他们。尽管学术体制为成功科学家提供了许多相信自己权威的理由，任何好科学家都明白，当你自信比最优秀的同学懂得更多时，你就不再是科学家了。

所以，科学群体既是一个道德规范群体，也是一个想象群体。

根据这个描述，我们清楚地看到争论对科学进步有着根本的意义。我的第一个原则说，当证据迫使我们达成共识时，我们就该达成共识。但我的第二个原则说，在证据达成共识之前，我们应该鼓励存在各种不同的观点。这对科学是有益的——费耶阿本德常这么说，我也相信是正确的。有理论竞争的时候，科学进步最快。过去有一种天真的观点认为，理论是在不同时间逐个提出来接受数据检验的。这没有考虑我们拥有的理论思想在多大程度上影响我们做哪些实验和怎么解释那些实验。如果一个时间只思考一个理论，我们很可能坠入那个理论产生的智力陷阱。唯一的出路在于让不同的理论竞争解释同一个证据。

费耶阿本德指出，即使有了一个广泛接受的满足所有证据的理论，仍然有必要提出竞争的理论来促进科学进步。这是因为竞争的理论最有可能提出与已确立的观点相矛盾的实验，而如果没有竞争的理论，那些实验甚至根本难以想象。因此，竞争理论不仅强化实验，也同样经常地提出反常实验。

于是，费耶阿本德坚持认为，科学家应该绝不认同，除非迫不得已。如果科学家在证据尚未确立之前太快达成一致，科学就危险了。于是我们要问，是什么影响他们做出不成熟的结论？他们也是人，人

们认同各种没有事实根据的事情，都是同样的因素造成的，如宗教信仰和大众文化的时尚潮流。

于是归结为这样一个问题：科学家达成一致，是因为他们想被其他科学家喜欢、赞美？还是因为他们认识的每个人都在想同样的事情？或者因为他们愿意站在胜利者的队伍里？多数人都是因为这些动机而忍不住赞同别人。没有理由要求科学家能例外，毕竟他们也是人啊。

然而，如果想保持科学的活力，我们还得与那些欲望进行斗争。我们必须鼓励反对者在实验允许的范围内有不同的观点。考虑到人人都渴望讨人欢喜，渴望加入成功者的队伍，我们必须清楚地认识到，如果我们屈从于这些欲望，就等于让科学堕落。

健康的科学群体应该鼓励分歧，还有其他的理由。当我们被迫认同某个意料之外的事情时，科学就前进了一步。如果我们认为我们知道答案，就会努力使每个结果去满足那个预先设定的思想。只有争议才能保持科学的活力，促进它不断进步。在观点激烈竞争的环境下，社会学力量不足以将人们驱赶到一个观点。所以，在某些少见的情形下我们确实为一件事情达成了共识，那只是因为我们别无选择。证据迫使我们不得不那样，尽管我们不喜欢。正因为如此，科学进步才是真实的。

科学的这个特征有几个明显的反例。首先，在我刚才说的规范群体中显然有叛逆者。科学家经常夸大和歪曲证据。年龄、现状、时尚、同行的压力等，都在科学共同体的活动中起着作用。有些研究项目能聚积超过证据所能支撑的人力和资源，而另一些最终结出丰硕成果的项目却在社会学力量的压迫下挣扎。

但我想指出的是，足够多的科学家在尽可能地维护着道德规范，因而从长远看来，科学还会持续地进步，尽管它浪费了一些时间和资

源来推进和维护原本错误的正统和流行的思想。我们必须强调时间的作用。不论短期内发生了什么事情，几十年后几乎总有足够的证据来解决争论，达成共识，而与时尚潮流无关。

另一点可能的反对意见是，我刚才说的特征在逻辑上不够完整。我没有拿出一个准则来说明哪些技术是必须掌握的。但是我想这最好由科学家群体经过很多代人的努力来决定。牛顿和达尔文不可能预言我们现在运用的工具和程序。

对一种规范的忠诚不会是彻底的，所以在科学实践中总还有改进的空间。这在时尚潮流（至少在物理学中）过分张扬的今天显得尤其正确。如今，你随处可以听到新获得博士学位的聪明的年轻人私下告诉你，他本愿做X，实际上却在做Y，因为那是有势力的老一辈们维护的方向或技术，他们只有跟着做，才好找工作、才能得到资助。当然，和在其他领域一样，科学中也总会有少数人选择做X，尽管他们很清楚做Y的人能在短期内得到更好的奖赏。在那些人中，很可能出现下一代的领导者。于是，科学的进步也许会因正统和时尚而缓慢，但从长远来看，做X的人还是有机会取代做Y的人，科学不可能完全停滞不前。

所有这些都说明，和人类从事的其他任何事情一样，科学的成功在很大程度上靠的是勇气和个性。虽然科学进步最终依赖于有多大的可能性在长远达成共识，但科学家个人决定做什么事情、怎么评价证据，却只能依靠不完整的信息。科学之所以进步，是基于在不完整信息面前人人平等的组织规范。没人能确定地预言一个方法是带来确定的进步还是浪费多年的劳动。我们所能做的只是训练学生掌握一些方法，经验已经证明那些方法能经常带来可靠的结果。然后，我们必须让他们去自由想象，必须花时间倾听他们的报告。只要科学共同体永远向新思想和新观点敞开大门，坚持道德规范，将最终的共识建立在根据公开证据进行的合理论证基础上，科学终将取得成功。

建立科学共同体的使命是永远不会完结的，它总是要和正统、时尚、年龄和地位等势力进行斗争。总会有挡不住的诱惑引导我们去走捷径，去迎合成功的团队，而不愿去认识新的问题。在理想的情况下，科学共同体应该让我们尽情发挥个人的冲动和激情。为了科学共同体的运行，必须将我们每个人都多少带着的傲慢和野心约束起来。还是费曼说得好：科学就是有组织地怀疑专家意见的可靠性。[144]

## 第十八章　预言家和工艺师

我们走的物理学革命路线也许有点儿问题。我在第十七章讲过，科学是人的行为，难免有人的弱点——也很脆弱，因为它需要个人的规范，也同样依赖于群体的规范。它也可能崩溃，我相信它现在就要崩溃了。

一个团体常常因为组织的原因而被迫以某种特殊的方式思考问题。一个重要的组织问题是：为了解决眼下的问题，我们是否组织并奖励了正确的问题和正确的物理学家？它对应的问题是：我们是否提出了恰当的问题？

任何关心基础物理学的人都会看到，新思想是必须的。从弦论的最大怀疑批评者到最热烈鼓吹者，你都能听到同一个声音：我们丢失了重要的东西。他们隐约感到需要某种新的东西，就是这种感觉促使2005年弦论年会的组织者们设立了一个“下一次超弦革命”的分会。尽管其他领域的实践者如今更有信心了，但我认识的每个物理学家都会赞同我们也许至少还缺少一个重要思想。

我们该如何寻找那个丢失的思想呢？显然，一定需要某个人站出来，要么找出我们大家现在都认可的某个错误假定，要么提出一个新问题。为了确保基础物理学的未来，我们正需要那样的人。这样，组织问题也就清楚了：我们是否有良好的体制以确保有人能在我们支持

和服从（这同样重要）的群体中间找出错误的假定或提出正确的问题？对这样有着罕见天才的创造性叛逆，我们是欢迎他还是驱逐他？

当然，善于提出真正新的然而又相关的问题的人是很难得的，而认清一个专业领域的现状，发现隐藏的假定或新的研究路线，更是一种特殊的才能，而且大不同于加入物理学群体所要求的那些基本技能。做一个技巧娴熟的手艺人是一回事，做一个有思想的预言家却是另一回事。

这种区别并不意味着预言家不是训练有素的科学家。预言家必须对学科有完全的了解，能用流行的工具进行工作，能用它的语言去说服别人。不过预言家不必是技术纯熟的物理学家。历史表明，与精通数学、善于解题的科学家比起来，成为预言家的那些人有时显得很平庸。爱因斯坦就是一个好例子，他年轻时连一个像样的科学工作也找不到。他与人辩论时有些迟钝，容易糊涂，而别人的数学都比他好。相传，爱因斯坦本人曾说过，“不是因为我太聪明，而是因为我能持久地考虑问题。”[145]玻尔是一个更极端的例子。历史学家贝勒（Mara Beller）曾详细研读过他的著作，指出在他的研究笔记里没有一个计算，尽是语言的论证和图画。[146]德布罗意（Louis de Broglie）曾提出一个惊人的建议，说假如光既是波也是粒子，那么电子和其他粒子也同样可以像波。这是他1924年在博士论文里提出的，当时并未引起考官们的注意，如果没有爱因斯坦的认可，他还差点儿不及格。据我所知，他再也没做过有同样影响的物理学工作。在我的想象中，只有一个人既有想象力也是他那时代最好的数学家：牛顿。其实，牛顿的每件事情几乎都是不可思议的。

第十七章说过，库恩区分了“常规科学”与科学革命。常规科学基于一定的范式，那是业已确立的关于固定理论、固定问题、固定实验方法和计算技术的实践。当范式被打破时，即当它所基于的理论不再能预言或解释实验结果时，科学革命就发生了。[147]我并不认为科学总是这样进行的，但一定有常规和革命的时期，科学在不同时期有不同

的做法。问题是，常规和革命时期需要不同类型的人。在常规时期，只需要能用专业技术好好工作的人，而不考虑他有多少想象力（当然也可能很高）——我们不妨称他们为工艺师。在革命时期，我们需要预言家，即那些能透过黑暗看清方向的人。

工艺师与预言家为着不同的理由来做科学。工艺师做科学，主要是因为他们在上学的时候发现他们很会做。他们从小学到中学一直到研究生院，通常都是数学和物理成绩最好的学生，然后走进同行的队伍。他们总能比其他同学更快更准确地解决数学难题，所以他们判断其他科学家就看是否会解题。

预言家就不同了。他们是梦想家。他们走进科学是因为想知道存在的本质是什么，那是课本没有回答的问题。如果他们不做科学家，就可能是艺术家或作家，或者一辈子待在神学院里。这样的两群人相互误会和不信任，当然是预料中的事情。

预言家通常抱怨物理学的标准教育忽略了科学发展的历史和哲学背景。有个年轻的物理学家想在他的物理学课中加入哲学，遭遇了很大阻力。爱因斯坦写信给他说：

我完全赞同你的意见，方法论与科学的历史和哲学有着同等的意义和教育价值。今天有很多人——还包括专业科学家——在我看来都只看见了千万棵大树，却从未看见整片森林。历史和哲学背景的知识能帮助人们摆脱多数科学家所沉迷的时代偏见。在我看来，哲学观带来的这种独立意识标志着普通工匠或技师与真理探索者之间的区别。[148]

当然，有的人是两者的混合体。没有高度专业技能的人是不可能将其坚持到研究生院的。但我认识的多数理论物理学家都属于两者之一。那么我自己呢？我想我是一个预言家，有幸也有一手好技能，还偶尔解决过某些问题。

我第一次看到库恩关于革命与常规科学的划分还在读大学，当时很糊涂，因为我说不清自己处在什么时期。如果考虑那些尚未解决的问题，我们显然正在经历着一场革命。但如果看看周围人们的工作，我们显然在做常规的事情。那时有一个范式，就是粒子物理学的标准模型和证明模型的一些实验，都是常规进行的。

现在我明白了，我的迷糊正预示着我在本书探讨的危机。其实我们就在革命时期，但我们想用过时的工具和常规科学的组织来摆脱它。

这也就是我对最近25年物理学的基本假定。我们正在革命时期，这大概是毫无疑问的。我们陷入了困惑，我们迫切需要真正的预言家。但预言家已经离我们很久远了。20世纪初我们有几个里程碑式的大思想家，首先是爱因斯坦，其次还有玻尔、薛定谔、海森伯等。他们未能完成他们开创的革命，但他们创立了部分成功的理论——广义相对论与量子力学，是我们继续革命的基础。这些理论的发展需要大量艰巨的技术工作，所以几代人的物理学都是“常规科学”，是工艺师们的天下。其实，20世纪40年代物理学的天下从欧洲向美国的转移，就是工艺师战胜预言家的结果。我前面讲过，它转变了理论物理学的风格，从爱因斯坦和他的伙伴们对基础的沉思，演变成为产生标准模型的激进的实用主义的态度。

当我在70年代学物理时，老师似乎都在教导我们俯视那些思考基础问题的人。当我们提出量子理论的基础问题时，我们听到的回答是，没人完全理解它们，它们关心的已经不再属于科学了。人们需要做的事情就是把量子力学当作确定的工具，将它用于新的问题。这是地道的实用主义态度；其格言是“少说话，多计算”。那些不甘心放弃对量子理论的意义的疑虑的人，被认为是不能做研究的失败者。

像我这样从读爱因斯坦的哲学沉思走进物理学的人，是不能接受那个理由的，但意思很清楚，我要尽最大的努力追随它。你可以在确

定的量子理论里经营自己的事业。普林斯顿高等研究院的幸福环境曾令我留恋，但那儿已经没有了爱因斯坦的科学作风的印迹——只有一尊空空的铜像在图书馆默默注视着外面。

但革命没有完成。粒子物理学的标准模型当然是实用主义物理学风格的胜利，但它的胜利如今似乎也标志着它的局限。标准模型（也许还有暴胀）大约是常规科学所能达到的极限了。从那以后，我们陷入了泥潭，因为我们需要的是回归革命的科学。我们再一次需要预言家。问题是我们周围几乎没有预言家，因为科学经过那么长久的常规研究，已经难得认识他们，更难容忍他们。

从20世纪初到70年代，科学（通常也包括科学院）越来越组织化和专业化。这意味着常规科学的实践被奉为好科学的唯一模式。即使人人都明白革命是必须的，我们群体中最强力的部分却忘了如何革命。我们一直在尝试着用最适合常规科学的研究风格和结构来发动革命。弦论的尴尬——承诺多而兑现少——恰好就是大量工艺师做预言家的工作所带来的结果。

我敢肯定有的弦理论家会反对这种说法。当然，他们在做物理学的基础问题，所有的工作都旨在发现新的法则。为什么弦理论家不是预言家呢？难道虫洞、高维空间和多重宇宙不是想象丰富的思想吗？是的，当然是，但问题不在这儿。问题在于：背景是什么？这些思想说的是什么？在卡鲁扎和克莱因思想经过3/4世纪之后，隐藏的维和虫洞一点儿也不新鲜了。在数以百计的人思考过同样的思想之后，考虑这些事情也用不着什么胆略或先见之明。

审视我们现状的另一种方式是，预言家为了满足他们清晰的渴望，被迫应对最深层的物理学的基础问题，包括量子力学的基础和与时空本性有关的问题。关于量子力学的基础，最近几十年里发表了很多论文和图书，但据我所知，没有一个作品是一流的弦理论家做的。我也不知道有哪个弦理论家写过什么论文，将弦理论面临的问题与物

理学家和哲学家关于空间、时间或量子理论基础问题的旧著作联系起来。

相反，量子引力的背景独立方法的倡导者们的科学观则是通过对基础问题的长期沉思而形成的。他们的思想产生了很多有关基础问题的论文甚至专著；也很容易罗列那群人物的名单：彭罗斯也许是公众最熟悉的一个。但我们还能列举别的人，如贝兹（John Baez）、克兰（Louis Crane）、德维特、多克（Fay Dowk-er）、伊沙姆（Christopher Isham）、马科普洛、罗维里、索金和特胡夫特。

相反，我想不出有哪位主流弦理论家提出过量子理论或时间本性的原创性思想。弦理论家往往以轻蔑的姿态来回应这些批评，大概会说这些问题都解决了。他们偶尔也承认问题是严肃的，但马上会接着声明现在还不是解决它们的时候。常听人说，我们应该继续紧跟弦理论的发展，因为弦理论是正确的，一定会包含那些问题的解答。

我绝不反对人们像工匠那样做科学，他们的工作需要坚实的技术做基础。正因为这样，常规科学才那么有力量。但指望能在现有理论之内通过解决技术问题来解决基础问题，简直是痴人说梦。假如真是那样就好了——当然，我们就会少想一点儿，即使对感到被迫那么做的人来说，思考真的很难。但深层的长久的问题从来不会靠偶然来解决，只有一心想着它们、决心直接攻克它们的人，才可能解决它们。这些人就是预言家，也因为这一点，科学机构欢迎他们而不是排斥他们，才显得那么重要。

科学从来就不是为了方便预言家而组织起来的；爱因斯坦的求职经历绝非个别例子。但在100年前，科学院还不那么专业，训练有素的外人也随处可见。这是19世纪的传奇，那时多数做科学的人都是狂热的业余爱好者，他们要么很富有，不需要做工挣钱，要么相信自己能找到赞助人。

那很好啊，你可能会说。但谁是预言家呢？他们本就是特立独行的人，把科学当生命，即使不能靠它生存也要做。尽管我们的专业研究机构对他们不好，仍然会有那么几个人。他们是准呢？他们打算做什么来解决那些大问题呢？

他们躲在我们视野之外。看他们对我们多数人相信的假定的拒绝，就可以发现他们的存在。让我来向大家引见几位吧。

要说狭义相对论是错的，我有很多疑惑。如果它错了，那么存在一种特殊的静止状态，能最终测定其方向和速度。但我们周围有少数理论家却不觉得有什么疑惑。雅各布森（Ted Jacobson）是我的朋友，与我合写过一篇圈量子引力的量子力学的论文。我们还一起发现了著名的惠勒-德维特方程的第一组精确解。[149]但正当圈量子引力汹涌向前时，雅各布森却悲观了。他认为圈量子引力不会成功，还认为它不够深刻。经过再三考虑，他开始怀疑相对性原理本身，相信有可能存在一个特殊的静止状态。他花了多年时间来发展这个思想。在第十三章和第十四章，我说过如果狭义相对论是错的，会很快有实验告诉我们结果。雅各布森和他在马里兰大学的学生就代表着一帮寻求狭义相对论实验检验的人。

另一个怀疑整个相对论框架的预言家是宇宙学家马盖若（见第十四章）。他别无选择，因为他发现并喜欢上了自己的一个似乎与相对论矛盾的思想——即光速在宇宙早期可能要快得多。他关于这个思想写的论文看起来很和谐——如果不假定要抛弃或至少修正相对论原理，它们当然也就没有什么意思了。

还有一个狂野的家伙，一个卓有成就的物理学家，是做固体物理学的，为解释材料的行为创造过辉煌的业绩。我说的是劳克林，1998年因为“发现新的具有很少荷电激发态的量子流体”的贡献获诺贝尔物理学奖，还有莫斯科朗道理论物理研究所的沃洛维克，他解释了一些极冷液氦的行为，还有MIT的文小刚。这些人既是工艺师也是预言

家。他们在最近几十年做了最好、影响最大的常规科学，然后又将手伸向量子引力的深层问题。他们的出发点是认为相对论原理错了，只不过是一种近似的突现现象。粒子物理学家贝约肯是另一个工艺师与预言家的复合体。我们如今知道质子和中微子包含着夸克，那主要就是靠他的洞察力。

有个大预言家叫涅尔森，来自玻尔研究所。他构建过弦理论，也有过很多大发现。但多年来他离开了主流，去宣扬他所谓的随机动力学。他相信，关于基本定律我们顶多只能假定它们都是随机的。我们认为本质上正确的每一件事情，如相对论和量子力学原理，在他看来都不过是从基础理论突现出来的偶然事件，而那基础理论远远超越了我们的想象，同样可以假定它的定律都是随机的。他的模型是热力学定律，我们过去认为那些定律是以原理为基础的，但现在我们将其理解为大量随机运动的原子的最可能的行为方式。这也许不对，但涅尔森在他的反统一纲领中已经走得很远了。

弦理论家中就没有多少人能像前面几位先生那样对科学有过持久的贡献。那么，当那些杰出的物理学家一再警告我们也许在做着错误的假定时，弦理论家——在这个问题上，或许也包括圈理论家——是如何回应的呢？我们才不管他们呢。是的，坦白地说，有时他们刚走出大门，我们就在背后嘲笑他们。虽然你做出了诺贝尔奖水平的物理，甚至已经得了奖，但如果你要质疑大家坚持的像狭义和广义相对论那样的假定，我们也是不会容忍的。我非常吃惊地听劳克林说起，他在系里和资助单位受到了很大的压力，人家要他做他原来的常规科学，而不愿他浪费时间做关于时间、空间和引力的新思想。像他那样卓有成就而且头顶诺贝尔桂冠的人都不能发展自己的深刻思想，还谈什么学术自由呢？

幸运的是，我们很快就会知道狭义相对论是否正确。我的多数朋友都希望实验观测能证明那些大人先生们都是傻瓜。我希望想打破偶像的人是错的，狭义相对论能经得住考验。但我也老是担心也许错的

是我们，而他们才是对的。

相对论的质疑就谈这些。如果量子论错了，情况又如何呢？这是整个量子引力计划的软肋。如果量子论错了，拿它来结合引力就纯粹是浪费时间。有人考虑过这个问题吗？

有的，其中一个就是特胡夫特。还在乌得勒支做研究生时，特胡夫特就和一个年长的合作者一起证明了量子的杨-米尔斯理论是合理的，这个发现使整个标准模型成为可能，他的这些成就当然赢得诺贝尔奖。那是他众多关于标准模型的基本发现之一。但最近10年他成了基础问题的最大胆的思想家。他的主要观点是所谓的全息原理。在他的构想里，没有空间。发生在我们常想象为空间的某个区域里的每个事件，都可以表示为发生在包围那个空间的曲面上。而且，描述那个边界世界的理论不是量子论，而是某个他相信能取代量子论的确定性理论。

就在特胡夫特构想他的原理之前，克兰在量子引力的背景独立方法的前提下提出了类似的思想。他提出，将量子论用于宇宙的正确方法不是将整个宇宙放进量子体系。霍金、哈特尔（James Hartle）等人尝试过那种方法，但遇到了严峻的问题。克兰反过来认为，量子理论不是一个系统的静态描述，而是一种信息的记录——它记录的是宇宙的一个子系统通过相互作用而获得的关于其他子系统的信息。接着他提出，将宇宙一分为二的每一种方式，都联系着一个量子力学的描述。这些量子态不是存在于这个或那个区域，而是存在于它们之间的边界。150

克兰的大胆建议从此成长为一类量子论方法，叫关系量子论，因为它们所依赖的思想认为量子力学是宇宙子系统之间的关系的描述。罗维里发展了这个思想，他证明它与我们通常的量子理论做法是完全一致的。在量子引力的背景下，它引出了一种新的量子宇宙方法，是由马科普洛和她的合作者提出的。马科普洛强调，描述不同子系统之

间的信息交换就等于描述决定一个系统影响另一个系统的因果结构。这样，她发现宇宙可以描述为一台能动态生成逻辑的量子计算机。[151]宇宙是量子计算机的思想也曾由MIT的劳埃德（Seth Lloyd）提出过，他是量子计算领域的蓝图设计者之一。[152]马科普洛和劳埃德从各自学科的两方面出发，发起了一场运动，用量子信息理论的思想来重构宇宙的概念，使我们认识了基本粒子是如何从量子时空突现出来的。

特胡夫特用边界来代表世界的思想，应该令我们想起马尔德希纳的猜想。其实，特胡夫特的思想正是马尔德希纳的灵感来源之一，还有人认为全息原理将成为弦论的一个基本原理。仅凭这一点，特胡夫特就很容易成为弦理论群体的领导者，假如他对这个角色感兴趣的话。但在20世纪80年代，特胡夫特开始走自己的路了。那时正当他的黄金年华，而且在技术上也没有人比他更强。不过，他脱离主流时，还是遭到了他的粒子物理学伙伴们的嘲笑。他似乎并不介意，甚至毫无察觉，但我相信那深深刺痛了他。不管怎么说，他几乎怀疑一切，独闯了一条自己的基础物理学道路。他几十年来形成的核心信念是，量子物理学是错误的。

没有谁比特胡夫特更认真和真诚的了。我们量子引力领域的人喜欢他的一点就是，到处能看到他的影子。他参加过我们的很多会，但他从来不在大厅里和其他大人物一起谈政治。他喜欢到每个分会场去，而一般只有年轻学生才会那样。每天他一大早就来到会场，穿着一尘不染的套装（其他人一般都穿牛仔裤和T恤衫），整天坐在前排，倾听每个学生和博士后的报告。他并不常发表评论，甚至偶尔还会打一两分钟瞌睡，但他对每个同行所表现的尊重，令人难忘。轮到他讲话时，他站起来，一点儿也不做作地讲他的思想和结论。他知道他独自走着一条路，如果他抱怨，我也不会惊奇。一个人怎么会放弃他应得的领导权杖呢？就因为他弄不清量子力学的意义吗？想想那到底说明了怎样的个性。

还有一个彭罗斯。简单地说，对我们理解和运用广义相对论，除了爱因斯坦而外，没有人比彭罗斯的贡献更多了。在我认识的不同领域的人物中，他是四五个最具天才和深刻原创力的思想家之一。他做过大数学和大物理。和特胡夫特一样，他在最近20年的许多工作也是基于他相信量子力学是错误的。他也和特胡夫特一样在构想取代它的东西。

多年来，彭罗斯一直强调，把引力纳入量子论将使它成为非线性的。这样就解决了测量问题，即量子引力效应导致量子态发生动力学坍缩。彭罗斯的建议在他的著作里有很好的描述，尽管还没有形成一个详尽的理论。不过，他和一些人已经在用它们来预言一些可以完成的实验，其中有的实验正在进行中。

我们有几个人认真研究了彭罗斯的论证，更有少数的人相信它们是对的。但多数弦理论家——当然都是些引领潮流的弦理论家——似乎根本没听进去。即使像这样的大思想家，当他们质疑基本假定时都遭人白眼，你就可以想象没有首先做出过重大贡献的预言家会有怎样的遭遇。153

如果当世的几个最优秀的理论物理学家觉得需要质疑相对论和量子论的基本假定，那么一定会有人从一开始就站在他们的立场。确实有些人在研究之初就在考虑量子论一定是错的。他们学会了它，和别人一样能运用它的论证和计算；但他们不相信它。为什么呢？

这样的人大概有两类：真的和假的。我属于那种从来不相信量子力学的人，但我也是假的。就是说，我在上学的时候就明白了，如果我一心想明白量子力学，那么我就不会有一个好的理论物理学家的前程。所以我决定做主流能理解和欣赏的东西，这样我就能找正常的工作了。

幸运的是，我找到了一个方法，可以通过做像量子引力那样的主

流工作来考察我对基础问题的怀疑。因为我一开始就不相信量子力学，所以很清楚我的努力注定会失败，但我希望能从失败中找到线索，看什么东西可以取代量子理论。如果早几年，我在量子引力的经历恐怕会和我在量子理论的经历一样倒霉。然而，当我读研究生时，有了一个好机会，就是用最近发展起来的研究标准模型的方法来解决量子引力问题。所以，我可以假装做一个常规科学的物理学家，像粒子物理学家一样训练。接着我将我学的东西用于量子引力。因为我是最先尝试这个方法的人之一，用的方法又是引领潮流的人所理解的，于是事业才有可能一帆风顺，当然还说不上如日中天。

但我绝不能完全压抑自己的本能，探索我的学科基础。我在1982年写过一篇题为“论量子与热涨落之间的关系”的文章，现在看它，简直不敢相信我有那么大的胆量。[154]我提出了一个新问题：如何将空间、时间和量子协调起来？这个问题开辟了一条全新的解决路线。即使今天我已经写过很多有影响的论文，还认为那篇是我最好的作品。偶尔我会遇到追溯学科基础的同学或在主流外徜徉多年的孤独者，他们会说，“呀，你就是那个斯莫林啊！我从没想到会见到你。我原以为你一定早死了，或者离开物理学了。”现在我终于和圆周研究所的同行们回到了原来的工作，研究量子力学的基础。

那么真正的怀疑者又如何呢？他们不相信相对论和量子论的基础假定，而且绝不改变自己的倾向。他们是一群特殊的人，每个人都有故事。

许多喜欢科学的人都知道巴伯（Julian Barbour）是《时间终结》的作者，他在那本书中说时间是错觉。[155]他是一个与众不同的物理学家，自1968年获科隆大学博士学位以来就没有找一个研究工作。但他在认真思考量子引力的小人群中影响很大，正是他教导我们那意味着需要一个背景独立的理论。

巴伯告诉我们，他在读研究生时，在一次登山旅行的路上，他突

然想起时间也许是一种错觉。于是他开始考察我们包含在广义相对论里的时间认识的根源。他意识到他疑虑时间本性就做不了传统的学术研究，他也明白如果他要继续研究那个问题，就必须全心投入，而不能受常规物理学的压力的干扰。所以他在距离牛津一个小时的小村庄买了一所旧农舍，还把新婚妻子带来，安心思考时间问题。大约10年后他才向他的同事们报告他的结果。在那期间，他和妻子生了四个小孩，靠他做翻译的钟点工抚育他们。他一周用来翻译的时间不到20个小时，于是有很多时间思考；这一点和学术机构的多数科学家一样，他们除了教书和行政事务外，也只有那么多时间来思考。

为了与广义相对论的时间意义发生联系，巴伯钻进了那门学科，回溯物理学和哲学的历史。最后他构想出一种新理论，其中的空间和时间只是一个关系系统。他关于这个主题的论文慢慢开始受到关注，最终成为量子引力群体的荣誉一员。他将爱因斯坦的广义相对论重新解释为一个关系理论，这也正是我们领域的人现在所理解的方式。

这还不是巴伯的全部工作，但已足以说明一个成功预言家的经历与传统的专业科学家有多么大的不同。这样的人不赶潮流——其实他们大概对学科还没有足够的认识，根本不知道它的潮流是什么。这样的人，他们的动力不是别的，就是以前形成的信念，确信别人遗失了某个关键的东西。他们的方法更有学术性，因为他们只有通读了困扰他们的问题的整个历史，才可能想清楚。他们的工作是高度专一的，不过需要很长的时间才能得到一点结果。不管什么结果都无助于他们的职业生涯。巴伯成熟了，比多数职业科学家更多地改变了科学；但当多数物理学家在谋求终身职位的时候，他却没有一样可以炫耀的东西。

巴伯的经历很像其他的预言家，如达尔文，他也是躲到英国乡间去找间屋子思考困扰他的问题。爱因斯坦花了10年的时间孕育狭义相对论的思想。于是，预言家要发现未经检验的假定，所需要的只是思考的时间和自由。剩下的就看他们自己了。

另一个例子是芬克尔斯坦（David　Finkelstein），乔治技术研究所的荣誉退休教授，一生都在追求自然的逻辑。他做物理的方式与众不同。我们第一次见面时，他告诉我他一生的工作就是追求理解，“上帝是如何构想出这个世界的？”除此而外他什么也不做。我们每次见面时他都有新的认识。这条路线也引出一些副产品。他是第一个认识黑洞事件视界的人。[156]他第一个发现了固体物理学的所谓拓扑守恒律的重要特征，也第一个研究了一系列不同的数学结构——如量子群。一个预言家在自己追求真理的道路上能有多大的成绩，他的一生树立了一个榜样。虽然芬克尔斯坦有职业，但如果一个人只听从学术圈内的声音而忽略外面的一切，还能像他那样在那个年代获得名牌大学的教授职位吗？做梦去吧。

还有一个故事，更像是巴伯了。瓦伦提尼（Antony　Valentini）从剑桥大学毕业（和巴伯一样），然后在欧洲漫游了几年，最后在意大利的里雅斯特定居下来，师从席艾马（Dennis　Sciama）——他在剑桥曾是霍金、彭罗斯、里斯（Martin　Rees）、艾利斯（George　Ellis）等大相对论专家和宇宙学家的老师。后来，席艾马移居的里雅斯特，在意大利一家新建的研究机构SISSA（国际高等研究院）成立了天体物理学小组。瓦伦提尼是席艾马的最后一个学生，并不做天体物理，而是做量子理论，他强烈感到量子论没有意义。他研究了德布罗意在20世纪20年代最早提出的一个老思想，叫隐变量理论。根据这个理论，量子论的方程背后存在着一个实在。隐变量的思想被压制了很多年——部分是因为冯诺伊曼在1932年发表了一个错误的证明，说那种理论不可能存在。50年代，量子理论家玻姆终于发现了那个错误，复活了德布罗意的理论。瓦伦提尼对隐变量理论做了新的重要修正，是那个理论在几十年来的第一次进步。他关于这个问题的多数论文都被物理学杂志拒绝了，但其内容如今在研究量子力学基础的专家圈子里广为流传。

席艾马尽可能地鼓励和帮助瓦伦提尼，但不论在意大利还是在英

语世界，都没有为专心做基础问题的人留下位置。席艾马劝他，如果不能在杂志上发表他越来越多的结果，可以写一本书出来。瓦伦提尼没找到工作，到了罗马，终于在罗马大学得到一个博士后的位置。离开学校后，他又在罗马呆了6年多，爱上了那座城市和那儿的一个人。他靠做家教谋生，同时发展自己的理论，将结果写成书。[157]

虽然许多一流的物理学家私下承认自己对量子力学的忧虑，但他们在公开场合却说它的问题早在20世纪20年代就解决了。后来有多少人做基础研究，没有专门的报告；但我知道，至少从50年代起，一流杂志发表这方面的论文就很挑剔了，还有几家杂志公开宣称不接受这类文章。资助机构和主要的政府基金一般都不支持这种研究，[158]大学物理系也不聘用做那种研究的人。

人们的普遍反对，部分是因为科学在40年代从革命转向了常规。正如政治斗争一样，为了巩固革命果实，必然会镇压反叛者。早年有几个相互竞争的诠释量子理论的思想。到40年代，其中的一个胜利了。为了捍卫玻尔的领导地位，人们称它为哥本哈根诠释。玻尔和他的追随者为了解决争论费了很大气力，我听说还动用了政治力量，这一点儿也不奇怪，因为他们卷入了核武器的研制，当然应该是胜利者。但即使不关心意识形态而只想安心做常规科学的人，也想把学科的争议平息下来。从实验和实践方面说，量子理论取得了伟大的成功，靠它进步的人不想被别人没完没了的怀疑所困扰，他们才不担心理论的建立和解释是否存在更深层的问题呢。现在是稳步向前的时候了。

坚定的怀疑者们几乎没有多少选择。有人学着哲学家的作风，在哲学杂志上发表长篇大论。他们形成了一个小小的文化圈，至少使争论延续了下来。几个有数学天赋的人在数学系找到了工作，发表正式的严格论证的、不同于量子力学共识的观点。有的人——学科内最优秀的人——在不知名的大学里做教授，在那些学校，用不着自己去找研究资助。还有几个人做了其他领域的物理，偶尔像业余爱好者那样

做量子力学。

在这些“业余爱好者”中，有个叫贝尔（John Stewart Bell）的，在60年代初发现了隐变量理论的一个关键定理。他靠粒子物理学的成就奠定了自己的学术生涯，但在他去世多年后的今天，人们才发现他最重要的贡献在于量子理论的研究。据说贝尔说过，一个人应该做常规科学，而只用十分之一的时间来关心量子理论。这句格言流传时，我在圆周研究所的同事哈代（Lucian Hardy）常常在想，如果贝尔在他影响最大的领域里多花点儿时间——当然，除非他想丢掉自己的饭碗——他的贡献该有多大呀。

量子力学基础的研究在这段时期几乎没有什么进展，也就不足为奇了。还能指望别的吗？当然，就凭这一点，即使有少数人取得了进步，也完全可能没工作，没资助，也发表不了文章。

现在我们知道怀疑者犯了多么大的错误。大约20年前，费曼和少数几个人就意识到，也许我们能以量子现象为基础制造一种新型的计算机。这个建议被长久地忽略了，直到1985年多伊奇（David Deutsch，现在牛津量子计算中心）才提出一个更详细的量子计算机计划。159没有像多伊奇那样的基础问题的思想家；他发愿做量子计算机是因为他对数学和量子理论的基础问题感到不安。想知道他是怎样一个独创和清澈的思想者，可以看看他那本刺激的书《实在的结构》，160他在书中精心编织了他的多世界理论。我一点儿也不赞同他的理论，但我喜欢那本书。

1994年，MIT的肖尔（Peter Shor，那时是贝尔实验室的计算机科学家）发现一个惊人的结果：足够大的量子计算机可以破解现有的任何密码。161从那以后，经费就滚滚流入量子计算机领域，因为政府是绝不想让它的密码被人破解的。这些钱培养了新一代年轻人，一群聪明的科学家——物理学家、计算家专家和数学家。他们开辟了一个新领域，融合了物理学与计算机科学，形成了重新检验量子力学基础

的重要力量。量子计算在一夜间火爆起来，涌现出大量的新思想和结果。有些结果触及了基础的角落，还有些结果是20世纪30年代后的任何时候都可能发现的。这个例子清楚地说明了科学政治对一个学科的压力如何阻止了它几十年的进步。

1999年，瓦伦提尼在罗马孤独地过了7年后，回到了伦敦的父母家。他们一家是从阿布鲁左的小村庄迁移来的；他们开过一家小商店，想靠它支撑他的工作。我做帝国学院访问教授时，在那儿见过他。和伊沙姆（他是那儿的理论组的头儿）讨论后，我们决定给他一个博士后的位置，把他带回了科学。我们能那么做，是因为我得到了一笔意外的慷慨资助，资助者正好很关心量子力学的基础问题。我觉得钱的最好作用就是用来资助有可能在领域取得新的重要贡献的少数几个人。如果我只有国家基金会（NSF）的资助，就不可能做到这一点。虽然NSF对我的量子引力工作很慷慨，但用那个资助来帮助一个博士后做量子论的基础工作，还是会影响以后的资助。

现在瓦伦提尼加入了我们的圆周。他还在写隐变量的书，同时也成了量子理论基础领域的头面人物，常去相关主题的会议做特邀报告。他经常发表文章，最近的研究涉及一个大胆的新建议：通过观测来自邻近黑洞的X射线检验量子力学。[162]他和巴伯一样，在几年孤独的日子里潜心自学，在整个量子理论领域找不出第二个像他那样有深刻洞察力和渊博知识的批评家了。

我们应该记住，如果巴伯和瓦伦提尼想挣一个普通的研究职位，就不可能有任何成就。如果做一个普通的助手或助理教授，就会为了赢得进阶所必须的邀请和资助而拼命去发表文章、争取荣誉，结果是一事无成。但巴伯和瓦伦提尼硕果累累。他们一直在思想，对一个顽固的问题会比助理教授们想得更深、更专注。当他们经过十年苦想破壁而出时，都会形成一个严谨、独创而成熟的观点，使他们很快产生影响。经过那些年的潜心钻研，他们得到了重要的新发现，树立了自己的权威，成为了关心那些问题的人们的核心。

在事业之初（甚至以后）忍受长期的孤独，是预言家们的基本经历。有人说格罗藤迪克（Alexander Grothendieck）是目前健在的最能干、最富想象的数学家。他有过最不寻常的经历。他具有重大影响的一些贡献都不曾公开发表，而是以数百页信件的形式寄给了他的朋友们，然后逐渐在能理解它们的小圈子里传播。他的父母为躲避政治迫害和战争而流亡他乡；他是在第二次世界大战后的难民营里长大的。他在巴黎数学界崭露头角时，就像从天上降下来的。短暂风光过后，他在20世纪70年代几乎脱离了科学生涯，至少部分是因为他反对数学为军事服务。1991年，他完全消失了，尽管谣传他在比利牛斯山隐居，但没人知道他究竟在什么地方。显然，他是一个极端的例子。但我们还是可以看到一些优秀的数学家在声名鹊起时表露在脸上的羡慕、惊奇甚至恐惧的表情。下面是他对他一些经历的描述：

我在那些动荡的年月学会了孤独。[但即使]这个形式也没能真正表达我的意思。不管怎么说我都不想学孤独，理由很简单：这类知识我在童年时就从没忘过。那是我们每个人与生俱来的基本能力。然而，在那孤独的三年[1945~1948]里，我沿着自己摸索出来的路线，靠着自己的能力忘我地工作，这为我带来了强烈的自信（不张扬然而持久），相信自己有能力做好数学，而与形成规则的任何共识无关，也与任何时尚无关……我说这些的意思是，为了把握我想学的东西，要靠自己的方式，而不要靠大大小小的派别——不论我本人所属的派别还是因为任何其他理由自诩为权威的派别——所认同的观念（公开的或默许的）。不管在中学还是大学，那些无声的共识让我明白，我们不该自寻烦恼去担心诸如“体积”之类的名词的真正意思是什么，哪个是“显然不证自明的”、“普遍知道的”、“毫无疑问的”……正是以这种“超脱”的姿态，才能成就自己而不是沦为共识的走卒，才能拒绝困守在别人划定的圈子里——正是在这种卓然不群的行动中才能发现真正的创造力。所有其他事情就是理所当然的了。

从那时起，我有机会在向我召唤的数学世界里认识很多人，既有

我的“长辈”，也有我这个年纪的年轻人，他们都远比我聪明，远比我有“天赋”。我羡慕他们的才能，运用新思想就像玩把戏，仿佛从摇篮里就开始熟悉它们了——而我自己却感觉笨拙甚至痴呆，痛苦地徜徉在崎岖的路上，就像一头笨牛面对一座望不到头的大山——那尽是我决心要学的东西，也是我觉得无法理解其本质的东西，无法追随到底的东西。其实，我本人几乎没有什么特质能算聪明学生，既不能赢得竞赛的名声，也不能轻松消化多数可怕的学问。

实际上，多数我判断比我聪明的同志都成了著名的数学家。不过，从30年或35年后的观点看，我可以说他们在我们今天的数学留下的印迹还不太深刻。他们都做过很多事情，通常还是很美妙的，不过那些都在他们之前就已经开始了，而他们也没想破坏它们。他们不知不觉地陷入了那些看不见的牢固的小圈子，将特定环境的世界划定在一个给定的区域。想要打破这些界线，他们必须重新发现自身的那种与生俱来的（和我一样的）能力：忍受孤独。163

2005年，人们常问年轻的爱因斯坦在今天会不会被大学聘用。答案显然是否定的；他在当时就没能进大学做老师。现在我们更加专业化了，聘用一个人要看他在经过高度专业训练的人中间的竞争力。我上面提到的那些人也都不可能被聘用。如果说我们能享有这些人的贡献，那是因为他们凭着自己的慷慨——或者说倔强——坚持不懈地奋斗，而不靠科学家们通常获得的科学世界的支持。

乍看起来这似乎很容易纠正。这种人不多，也不难识别。很少有科学家思考基础性问题，而能提出重要思想的就更少了。我的朋友考夫曼曾告诉我，不难发现那些有大胆思想的人——他们差不多都至少有几个那样的思想。如果他们在研究生毕业时还没有什么想法，也许就永远不会有了。那么，该如何区分有好想法的预言家与那些正在努力却没有想法的人呢？这也容易，只要问问老一辈的预言家。在圆周所，我们可以毫不困难地发现几个值得关注的年轻人。

但发现这些人后，就需要将他们与其他做常规科学的人区别对待。他们多数都不关心谁更聪明或谁能更快解决主流的常规科学问题。如果他们要去竞争，即使在一个严格的小圈子里，他们也会失败。如果说他们在和什么人竞争的话，那就是最近的一代革命者，那些人在别人已经不看的旧书和论文里与他们对话。驱动他们的几乎没有什么外在的力量；他们只关注多数科学家视而不见的科学中的矛盾和问题。你等上五年甚至十年后，他们照通常标准也不见得好。但不要惊慌，就让他们自己去想吧。最后，他们会像巴伯和瓦伦提尼那样脱颖而出，证明他们是值得等待的。

那样的人不多，因此在科学机构里为他们留下一些空间也不难。实际上，我们应该想许多科研机构和大专院校都会为拥有那样的人感到幸福。因为他们对本学科的基础问题看得很清楚，所以也通常是优秀的甚至有超凡魅力的老师。没有什么能比预言家更能点燃学生的思想火花。因为他们不讲竞争，所以是很好的导师和引路人。大学的主要任务不就是教书育人吗？

当然，也存在风险。他们有些人可能不会发现任何东西。我这是从人的一生对科学的贡献来说的。但多数研究型的科学家，尽管在职业生涯中成功了——得到了基金，发表了很多论文，参加了很多会议，等——对科学的贡献却只是增加了一点东西。我们理论物理学的同行里，至少一半的人没能做出独特的真正持久的贡献。做好本职工作与做基础工作是完全不同的经历。但是，如果要他们一生做别的事情，那么科学就会沿老路走下去。所以，这又是另一种风险。

不同风险有不同的性质和代价，生意人比研究机构的管理者更懂得这一点。很容易与生意人谈论这个问题，但学界人物就不好说了。我曾向一个成功的风险投资者打听，他的公司如何确定风险大小。他说如果他投资的公司有10%能赚钱，他就知道风险不大了。这些人所理解和伴随一生的信念就是，即使90%的新公司都破产了，你也算获得了最大的回报，这也对应着技术进步的最大速度。

我希望能与国家科学基金会就风险问题展开真诚的对话。因为我相信，以实际的标准来衡量，他们资助我们领域的90%的项目都失败了。这些资助是不是促进了新的科学呢？——所谓新科学是说，如果受资助者不做那些事情，就不会有它的出现。

精明的商人都明白，低风险-低回报与高风险-高回报策略是有区别的，它们从一开始就有不同的预期目标。如果你想经营一家航空公司、一个公交系统或一家肥皂工厂，那么你就选择前者。如果你想开拓新的技术，那么没有第二种策略就不可能成功。

我不是要大学的管理者也像这样思考问题。他们在确定选择和提拔人物的标准时，仿佛只想到了常规科学家。其实，稍微改变这些标准以适应不同类型、不同才能的科学家，应该是再简单不过的事情。你想科学发生革命吗？那就学学生意人在期待技术革命时做的事情吧：改变一点法则，让几个革命者进来。不要把等级搞得那么森严，要给年轻人更多的空间和自由。平衡在低风险的科学累积上的巨大投入，为高风险-高回报的人们创造机会。这就是技术公司和投资银行采取的策略。科学团体为什么不尝试一下呢？我们的回报将是发现宇宙运行的秘密。

## 第十九章　科学到底是怎么运行的

在大学里改变科学的研究方式的想法，无疑会打动某些人，也会吓倒另一些人。不过那也许不会发生。为了解释为什么，我们需要考察学术生命的潜藏软肋。因为，正如社会学家告诉我们的，那不仅是智慧的事情，也关乎权力：谁拥有权力，在怎样使用权力。

研究型大学存在一定的阻碍改变的特征。首先是同行评议的制度，即让一些科学家对另一些做出评判。就像终身职位一样，同行评议带来了好处，所以人们才普遍相信它对好的科学实践有着根本性的

意义。但它也付出了很多代价，我们必须了解它们。

我确信一般人不会知道科学家要花多少时间来决定对其他科学家的聘用。我在委员会里每个星期大约要用五个小时来决定其他人的命运，或者给类似的委员会写信。我已经做了一些时问，那是一个教授的主要工作的一部分，我认识的许多教授在上面花的时间更多。有一件事是肯定的：除非你故意糟蹋自己，让大家看见你不负责任，或让人觉得你捉摸不定，没有诚信，那么，你做科学家的时间越久，你干预其他科学家经历的时间就越多。那不仅是因为你有越来越多的学生、博士后和合作者需要你为他们写推荐信，你还得替其他大学和机构决定人员录用。

有哪个管理者研究过这样的体制耗费了我们多少心血吗？我们能为它少花点儿时间而将更多时间用于科学和教学吗？我对这种体制没多大好感，它令人沮丧。任何一个有抱负的机构在聘人之前都会向由其他机构的有影响的老科学家组成的访问和咨询委员会网络征求意见。在美国、加拿大、欧洲和世界各地，资助机构还专门成立了评估小组。于是，为了公正而秘密地评价候选人，可以采取各种方式，如非正式接触、电话咨询、正式谈话等。经过一定时间以后，一个成功的科学家很可能要把所有时间花在人事决策上面。

这就是所谓的同行评议。名称很可笑，因为它明显不同于同行陪审团的概念，那意味着你将由一群像你一样的被认为是客观公正的人来评判。对那些暗藏偏见的陪审员，有真正的处罚——进监狱。

在学术界，评价你的人几乎毫无例外都比你老，比你有权力。从你大学第一堂课开始到你做教授申请基金资助，都是如此，我不想贬损那么多同行评议人所做的艰苦工作，多数都是真诚的。但里面存在一些大问题，关乎物理学在今天的状况。

同行评议在无意中产生一个负面结果：它很可能成为一种老科学

家强迫年轻科学家的机制。这种事情太明显不过了，可我奇怪为什么很少有人讨论它。体制的建立是为了让我们这些老科学家能给有前途的年轻人提供好机会，而将没有资质的人赶出科学共同体。如果有明确的标准和方法以确保我们客观公正，那当然是好的。可至少在我工作的部分机构，这是做不到的。

我们已经详细讨论过，不同类型的科学家在为理论物理学做着贡献，他们有不同的优势和弱点。然而，很少有人承认这一点；相反，我们只是简单说谁“好”谁“不好”——就是说，同行评议就基于这样一个简单而显然错误的假定：科学家可以像楼梯那样分出等级。

1984年，我在耶鲁大学当助理教授，当我看到第一批推荐信时简直不敢相信。一封好信可能会包含大量信息，而且尽量写得很委婉，大量心思都放在最后一段，通常是对候选人进行比较：“此人比A、B和C好，但不如E、F、G和H。”现在我已经读过几千封推荐信了，至少一半都有那种句子。那些年里，有几次我成了ABC。我同意“此人”比我好——事实上他们也做得很好。但如果研究表明这些排序多半不能预言真正的科学成功，我也毫不奇怪。如果我们真想用好人，就必须开展这种研究。自然有不少例子，拔尖的博士后或助理教授没能做多少事情，也没获得终身职位。

更大的问题还在于对偏见没有惩戒的措施。一个教授写信偏袒他自己的学生或追随他的研究纲领的人甚至他本国的人，一点儿也不觉得羞耻。我们可以看出（或笑话）那种露骨的夸张，但没人觉得稀罕。那不过是体制的一部分。

老科学家愿意推荐一种鉴别年轻科学家的基本法则，就是看那个年轻人会不会令他们想起当年的自己。如果在某人身上看到了自己年轻的影子，那人一定就是好的。坦白地说，我知道我为此感到内疚。如果你想聘用更多像我这样的人，我会把他们选出来。如果你想在和我迥然不同的人中进行选择，而他们做的事情我又不在行，那就不要

相信我的判断。164

即使对我们这些想做到公正的人来说，也没有什么指南或训练教我们如何公正。从来没人忠告我应该如何写或解读推荐信，我也没见过有什么指南告诉我们如何判别本人或他人意见中的偏见或敷衍。我在许多招聘委员会做过事，但从来没人像告诫年轻人那样指导我如何评价相关的材料。

在一次晚宴上，我问其他行业的人是否有过这类训练。那些不做研究但负责聘用或指导别人的人，都参加过几天培训，学习如何消除等级影响、鼓励思想的独立性和多样性、认识和避免偏见和不公正。他们都明白“注意倾听组织内的每一个声音”，从候选者指导过的人和指导过候选者的人那儿寻求评价，从而“360°地审视候选者”。如果说律师、银行业者、电视制作人和报刊编辑都需要人事决策的指南，为什么科学家就自以为能自动做好呢？

还有更糟糕的呢。在正式推荐信背后，还要和专家进行一系列私下的非正式谈话：“你对某某的看法如何？你认为我们应该录用谁？”

这些谈话很坦率，没有一点儿客套。这并不算坏事。很多人想揽些事情，显得乐于助人，但多半不讲什么客观。特别是，在这种情形，如果你想利用这种体制帮助你的朋友和朋友的学生，不需要付一点儿代价。成名的专家推举他们自己的学生和博士后，无原则地夸奖他们，贬低他人（特别是对手的学生），已经是司空见惯的事情了。

即使在这些坦白的交易中，也很少听到真正的反面意见。当人们没有好消息报告时，常常就说，“让我们继续吧。我没有意见。”或者说些无关痛痒的话，“我没有激动。”但有时候也会有人说某某“绝对不行！”“别那么说”，“你没开玩笑吧？”或斩钉截铁地说，“除非我死了！”据我的经历，出现这种情况，往往是因为候选者属于下面3种情形之一（通常是其中的两种）：①女性；②不是白人；③不随主流而

做自己的研究。当然，有时女性和黑人也没遭人反对。但在我的经历中，这些候选者都是死抱着既有的研究纲领的人。

物理学家们在激烈争论，为什么在物理学中妇女和黑人不像在其他同样有挑战性的领域（如数学或天文学）那么多。我相信答案很简单：公然的偏见。任何人，如果和我一样在招聘委员会服务过多年，却看不见行为中的赤裸裸的偏见，那他要么是瞎子，要么是骗子。因为要替人保密，我不能举具体的例子，但有几个详细的研究报道过这样的故事。165

也许我们本该想到这个领域的偏见有那么强烈。很多一流的理论物理学家过去不也是长着粉刺的懵懂小男孩儿吗？他们不是在他们得意的战场（数学课）上报复过那些夺走女孩儿的四肢发达的家伙吗？我也算其中之一，总要做点儿事情出来，至少要让那些家伙们吃点儿苦头——那完全是为了自信。不过我还记得对自己的代数能力颇为得意，我敢说，至少我本人深信数学技能是男人的天性。可为什么有那么多女性没有多大困难就能聘为纯数学家呢？因为如果你在数学中做出了优秀成绩，大家都能看得清楚。定理要么证明，要么没有证明；而理论物理学家高低的评判却十分模糊，随处都有偏见。例如，要区分好的理论家与自负的理论家，就不是轻而易举的事情。还注意一点，虽然天才的女音乐家历来就有，但是，只有当乐团让候选者隔着屏幕演奏时，录用的女性人数才大幅度增长起来。

于是才有了所谓“平权运动”。在我的所有经历中，我从没看见哪个妇女或黑人是通过平权行动计划被聘用的，他们并不需要——就是说，他们已是不容争议的最好申请者。当招聘委员会不完全由白人组成时，就没有人指责公然的偏见了，于是我们可以放宽平权行动。事实上，与众不同的人——因为这样那样的原因做出了令男性老物理学家不满的事情——是不会被录用的。平权行动是为了外表不同的人，如黑人或女人。但对那些只是思想不同的人——拒绝主流方法而偏爱自己思想的人呢？是不是也应该给他们平权的机会呢？

我们许多参与同行评议的人都怀着美好的愿望客观公正地选拔人才。当其他条件都相同时，我们选择更值得的候选者。就是说，在年龄和背景相当的、做相同研究项目的白人中间比较，我们的体制一般会选择更聪明、更用功的人。但问题是，在达到“同等条件”之前，我们必须做很多选择。在决定之前，那些选择的过程都是政策性的。这是有权势的老科学家向年轻科学家施加影响的基本机制。

这就导致强力舆论的形成，老科学家通过它来确保年轻人走他们的路。发挥这种权力有几个简单方法。例如，教师职位的候选者需要很多人（都比他更有权威）为他写推荐信。如果有一封评价不那么好的信，就可能令他失去机会。当我第一次面对那么多推荐信的时候，简直不知所措。其实，从三四封信就肯定可以看清候选者的情况。可为什么名牌大学却要十封、十五封呢？

一个理由就是，目的不仅仅是要聘用好科学家。招聘委员会、教授、系主任心里都装着别的目的，就是提升本系的地位（其实能保住现有地位就算幸运了）。我说的是比青年科学家的前程更有分量的东西，因为一个系的地位靠的是很多排名，而排名却由外人根据基金资助和论文引用数量的印象来评估。系主任和教授们不得不关心这个，因为这些事情直接影响经费的多少，关系到他们自己作为管理者的前程。于是，最重要的是聘用可能赢得大量资助的人。这对那些做大项目的人有直接的好处，但不利于想启动新项目的人。要来更多的推荐信，你可以看出举足轻重的老科学家是如何评价候选人的。于是，目的不是要找人来做好科学，而是要他来在短期内提升本系的地位。难怪招聘委员会不会为长远的问题伤脑筋，不会去考虑候选人是不是有创造性的思想，能在20年后产生影响；他们只需要知道有十几个老科学家看好某个候选人，说他是他们群体的拔尖人才。

但是，为了得到那么多信，你必须参与重大研究项目。如果你的项目很小，能评价你的高层人物还不到10个，那么你就只好请别人来评价，他们也许不赞同你做的东西，或者他们的项目正和你的在竞

争。于是，只有靠人数来说话了。显然，大研究项目占尽了优势！

这种机制无疑便宜了弦理论，而要探索其他研究纲领的人就难了。最近，《纽约时报》有篇文章说，“科学家们还需要发展更完整的东西，他们原来以为那些碎片最终能成为一个完整的理论。不过，弦理论家们正在接管通常属于成功的实验家们的果子，包括联邦的资助、崇高的奖赏，还有大学的终身教授位置。”在同一篇文章里，还引用了格罗斯（现在是圣巴巴拉加州大学卡夫利理论物理研究所所长）的一句话：“如今，如果你是一个不错的青年弦理论家，那你算是赶上了。”[166]

我不是要批评弦理论；弦理论家的行为不过是任何主流研究项目的做法。问题是我们的决策体制太容易受某些激进研究项目的摆布，却不管它的结果如何。同样的体制也曾一度妨碍过弦理论家。正如记者陶贝斯（Gary Taubes）讲的：

1985年8月4日，我和德鲁约拉（Alvaro de Rujula）在CERN的小酒吧喝啤酒……他预测90%的理论家会做超弦及其与超对称的联系，因为那是很时髦的。当他明确表示这种情况不妙时，我问他愿意做什么。他没有直接回答，把话题岔开了。“必须记住，”他告诉我，“对超弦理论的发展最有责任的两个人，即格林和施瓦兹，曾花了10～15年的时间系统研究当时并不时髦的东西。实际上，人们都在嘲笑他们的固执己见。所以，当人们想说服你应该做最时髦的课题时，别忘了最大的进步往往是那些不做最时髦的事情的人取得的。”[167]

我和一个名牌大学的系主任讨论过这种状况，他很懊悔在20世纪80年代初没有说服同行录用施瓦兹。“他们认同他是个非常聪明的理论家，”他说，“但我没能说服他们，因为他们说他太痴迷，除了弦论什么也做不了。而现在，我说服不了我的同事们录用任何不做弦论的人。”

我还想起和派斯（Abraham Pais）讨论过这些问题，他是粒子物理学家，也是爱因斯坦和玻尔的传记作者。我们在纽约洛克菲勒大学时常在一起午餐，他是那儿的教授，而我也在那儿工作过。“你也无能为力，”派斯告诉我，“我那时也这样，他们都是婊子养的！”

我想派斯扯远了。我没有说人，而是说我们该怎样构建科学的决策，保证促进科学进步所需要的那些科学家能有用武之地。

这种体制对物理学危机还有另一点重要影响：技术娴熟而缺少思想的人比有自己思想的人更有机会，因为简直没有办法衡量那些独立思考的人。这种体制的建立不仅是为了做常规科学，也为了确保常规科学就是大家做的。我研究生毕业申请第一个工作时，就明白这一点了。一天，我们正在等申请结果，一个朋友走过来，很焦虑的样子。原来是一个老同事要他告诉我，我可能没机会得到任何工作，因为我无法和别人比。如果我想工作的话，就必须停止自己的思想，做别人做的事情，因为只有那样他们才能拿我与同行比较。

我不记得我是怎么想的，也不知道为什么它竟然没让我发疯。我比别人多等了两个月才找到工作，这并不可笑。我已经想好了自己要做什么，那用不了我多少研究时间。但后来我走运了。圣巴巴拉理论物理研究所刚成立，有一个量子引力计划，于是我的生涯没有到头。

但我现在才明白到底发生了什么。没有谁在故意刁难我。我朋友和他的导师都是在替我着想。但这不过是社会学家想象的。原来，让我朋友传递消息的那位老同事新开了一个研究项目，需要很艰难的计算。那个项目需要聪明而敏捷的年轻人。他们告诉我，如果我做他的项目，他就给我一份工作。这是世界上最简单古老的交易：工人靠他的劳动挣一碗饭。

交易的方法有很多，接受的人受奖励，反叛的人——偏爱自己的思想而不顾老一辈思想的人——受惩罚。我的朋友卡洛·罗维里想在罗

马工作。有人让他去找某某教授，那教授很友好，向卡洛解释了他和他的小组正在进行的整个研究计划。卡洛很感谢教授的解说，也向他讲述了自己的研究计划。谈话很快结束了，卡洛没有得到希望的工作。我只好向他解释事情的原委。我们都一样太天真了，竟以为人们会奖赏那些有自己思想的人。

实际上，卡洛在成为他那领域的欧洲一流科学家之后，才得到罗马大学的工作。当他在其他地方做出了有影响的业绩，当世界上成百上千的人开始研究他的思想——只有这个时候，罗马的大教授们才会来听他讲他的思想，而那些思想是他刚做博士的时候就想带给他们的。

你可能想知道，卡洛是怎么得到第一份工作的。我这就告诉你。那是20世纪80年代后期，广义相对论领域由几个老人主持，他们曾是爱因斯坦的学生，而且坚持认为应该鼓励有最好、最独立思想的年轻人。他们领导着一个当时所谓的相对论群体，研究小组分散在美国的十几所大学。他们在那个领域几乎不再领头了，但还把持着少数职位，大概每两三年有一个新教师职位的空缺。卡洛是罗马的博士后，但由于官僚体制问题，他的工作从来没有被正式认可，而他也没领过一分钱的薪水。每个月人家都告诉他还要等开会研究，有了正式文件以后，他就能拿到支票了。就这样过了一年半，他打电话给美国朋友说，虽然他不想离开意大利，但他受够了。他问美国是不是有什么工作。碰巧，一个相对论中心正在找助教，他们听说卡洛可能会申请，就瞒着他在一周内把职位的事情办妥了。需要说明的是，那个中心没有做量子引力的——他们用卡洛是因为他在那个领域有独创而重要的思想。

这样的事情今天会发生吗？不大可能，因为现在整个相对论领域也被老科学家确立的一套程式化的大研究项目占据了。这与实验的引力波天文学有关，他们还希望（多年后仍然希望）能通过计算机的计算来预言实验能看到什么。如今，年轻的广义相对论专家或量子引力

专家如果不解决这些问题，就不大可能在美国的任何地方找到工作。

不管什么领域，只要尝到了成功的滋味，就能使从前的叛逆者转变为他们研究计划的捍卫者。在我自己的量子引力领域，我曾不止一次地被人中断，为的是支持聘用其他领域的有新思想的人，而不用在促进现有研究的狭小问题上用功的技术娴熟的人。

这里存在两个问题，有必要将其区别开来。一是老科学家决策的主导作用，他们经常动用他们的权利支持人们做他们在想象力丰富的年轻时代设计的研究计划。二是令各大学感兴趣并愿意聘用的科学家类型。他们是愿意请人来做某个特殊领域的每个人都能理解和判断的工作，还是请人来开辟他们自己的难以被大家把握的方向？

这牵涉到风险问题。优秀的科学家常常在同行中引起两种反响。常规的低风险的科学家一般会引出同一种声音，每个人对他们的感觉都一样。高风险的科学家和幻想家通常会激起两种极端的反响。有些人相信他们，热烈响应他们；其他人则尖锐地批评他们。

学生评价老师也会发生同样的事情。有某个类型的优秀老师，学生们并不觉得他公正。喜欢他的人会说，“这才是最好的老师！我以前从没见过；我就是因为他才上大学的。”但另外的人却毫不掩饰地在评价表上表达他们的愤怒和怨恨。如果你把评分都平均了——像决定教授加薪或升职时打分一样，将学生的分数约化成一个数字——就将错失这个关键的事实。

这些年来，我注意到反响的极端分布预示着一个科学家未来的成功和影响。如果一些人认为某某代表了科学的未来而另一些人认为他将带来灾难，可能就意味着那人确实是个了不起的人物，会一往无前地推动他自己的思想，并有能力和毅力坚持到底。这样的人需要一个能包容冒险者的环境，而反对冒险的环境是不会让他们走进去的。

就美国的研究型大学来说，基本事实是评价意见两极分化的人通常找不到工作。尽管我只是在自己的领域看到这种现象，但它可能是普遍性的。看看下面的科学家吧，他们都因为对进化论认识的大胆而有创造性的贡献赢得了大家的景仰：巴克（Per Bak）、考夫曼、马古利斯、帕朱斯基（Maya Paczuski）、特里弗（Robert Trivers）。其中两个是数学家，研究自然选择的数学模型，其他两个是著名的进化论学家。他们没有一个在顶尖大学里工作。我年轻时常感到惊讶。后来我才意识到，他们的理智太独立了，而他们的形象犹如两个独立的高峰：如果说有很多人仰慕他们，也同样有很多有权势的学者怀疑他们。实际上，如果以通常的优秀科学家的判断标准来衡量，这些创立新思想的人往往是有缺陷的。他们也许太鲁莽，也许不注意细节，也许在技术上有欠缺。这些批评通常适用于原创性的思想家，是好奇和独立将他们引进了一个他们未曾经过训练的领域。不论他们的见解多么新奇和有用，在本领域的专家们看来，他们的工作在技术上是不能令人信服的。

确实，这些独创性的科学家中，有的不太容易共事。他们没耐心，如果不赞同你的意见，就会直截了当表达出来；他们也缺乏优雅的举止，把正确看得很重，而不在乎与别人和谐相处。我认识几个这样的“困难人”，我怀疑他们愤怒的原因和科学界的聪明女人一样：一生都痛苦地觉得自己是边缘人。

这类问题当然影响了巴克的生涯，几年前他悲剧性地死于癌症，才54岁。他有着罕见的才能，可以在他专业以外的几个领域写论文，涉及经济学、宇宙学和生物学。那本该成为他骄傲的资本，能去最好的大学，但事与愿违，因为他毫不犹豫地指出他的解决方法可以引出专家们错过了的结果。如果他将创造力用于一个领域，他本该拥有更辉煌的生涯，但那他就不是巴克了。

你大概惊讶，为什么那些做部门领导的聪明人没有认识这一点，让那些人为他们的大学发挥作用。当然，有人意识到了，也愿意聘用

这样的人。最近几十年来，美国出现了向量子引力的非弦论方法开放的职位，不过大多数那样的职位都是因为那个领域在学校不出名，领导者很难请来名教授。于是，他从通常的政策解脱出来，通过成本效益计算，相信用一个不受重视的领域的人，会很快赢得一个顶尖的团队，提高本部门的地位。

实际上，我们讨论的问题也影响着所有科学，其他领域的几个有影响的老科学家表达了他们的担心。阿尔伯茨（Bruce Al-berts）是生物学家，曾任美国最权威、最有影响力的科学家组织美国国家科学院院长。2003年4月，他在就职讲话中指出：

我们建立了一个青年科学家的激励机制，但它太缺乏冒险精神。在很多方面，我们是自己最大的敌人。我们为了审查资助申请而建立的学术部门是由声称尊重科学冒险的同行组成的，但在分配资源时，他们一般都支持没有风险的科学。这对创新产生了巨大的削弱效应，因为我们的研究大学在选择新老师时就找那些能获得资助的助教。这就解释了为什么那么多的好青年在做着“应声虫”的科学。

接着他讲了一种趋势，在1991年以来的10多年里，国家卫生研究院（NIH）给35岁以下的年轻人的资助减少了一半，而给55岁以上的人却增加了50%。他为这个结果感到悲哀，因为它极大损害了年轻研究者的思想独立：

我的许多同事和我都是在30岁以前就得到了第一个独立资助。我们那时还没有初步的结果，因为我们做的是全新的东西。[现在]几乎没有一个人能在35岁以前开始自己的独立科学生涯。而且，1991年，NIH资助的35岁以下的主要研究者占三分之一，而到2002年，这个比例减小到了六分之一。哪怕我们当中最有才能的年轻人，也不得不遭遇申请被拒绝的命运，等到几年之后他们有了足够的“初步数据”，才能使评论者们放心他们可能实现他们提出的目标。

如果说问题那么明显，惊动了美国科学最权威的领导者，为什么没有什么作为呢？这令我困惑了很长时间。现在我明白了，获得并保持学术地位的竞争并不仅仅在于业绩。我们的体制试图选择最优秀、最有成果的人，在一定程度上也做到了。但还有其他的程序，忽略它们未免过于天真。同样重要的是，这些决定是为了在每个领域内部达成并强化共识。

用人不是达成共识的唯一渠道。我说过的关于用人的每件事情，也适用于评价资助申请的专家小组，同样还适用于职位的评估。这些事情都是相互关联的，因为如果你不能成功获得资助，就不可能在美国研究大学得到科学研究的职位；只有当你可能得到资助时，才会有人聘用你。

我第一次指出这一点时，碰巧有人请我为《高等教育纪事》（大学管理者的一种行业杂志）写一篇其他主题的文章。我写信给编辑建议另写一个题目，谈普通研究项目的主导作用对学术自由的威胁。他们很乐意看看，但看了我的草稿后就拒绝了。我感觉被侮辱了：他们在压制不同的声音！于是我给他们发了一个极不愉快（对我来说）的电子邮件，质问他们的决定。他们很快就回信了，告诉我问题不在于文章太激进——恰恰相反。里面的每件事情都是众所周知的，在社会科学和人文科学中已经广泛讨论过了。他们给我寄来一堆他们在过去发表的讨论学术决策中的权利关系的文章。我通读了一遍，才发现只有科学家对这些问题视而不见。

显然，拥有长期的职位是很有道理的。在有限的程度上，它保护了有创造力的独立的科学家不会被追赶智力时尚的年轻人所取代。但我们也为终身教授的制度付出了沉重的代价：对年纪较大的人来说，工作太保险，权力太大，而责任太少；对最富创造力和冒险精神的年轻人来说，工作太没有保障，权力太小，而责任太大。

尽管终身职位保护了智力独立的人，却不能创造那样的人。我听

许多同行说过，他们做时髦的事情是为了得到终身职位，然后才做他们真正想做的事情。但事实似乎并非如此。我只知道一个那样的例子。而在其他情形，如果那些人在为职位焦虑时没有足够的勇气和独立去做想做的事情，那么他们在考虑专家组如何评价他们的资助申请时，也不会突然勇敢和独立起来。如果一个体制不能让思想独立的人得到终身职位，那么同样的体制也不会有助于保护终身教授的思想独立。

实际上，有的终身教授因为转向更冒险的领域而失去了资助，他们会很快发现自己陷于困境。他们不会被解职，但也有各种威胁的压力。他们可能被加重教学负担，削减薪水。于是，他们要么回到低风险的研究，要么提前退休。

下面是MIT天才的数学教授辛格（Isador　Singer）最近对他的学科的描述：

我发现在经济因素的驱动下有过早专业化的趋势。你必须很早表现出大有前途才能得到好的推荐信，换来好的工作。在你崭露头角、巩固地位之前，根本没有能力另辟蹊径。生活的现实强迫你用狭隘的观点来看数学，而那是与数学不相容的。我们可以利用新的资源来对抗过分的专业化，给青年人更多的自由，让他们更大胆地自由探索数学，探索数学与其他学科的关联——例如在生物学中，当前发现了很多东西。

我年轻的时候，工作的市场还不错。在主流大学工作当然是重要的，但在小一些的学校同样可能成功。我为时下工作市场的强迫效应感到痛苦。年轻的数学家应该有我们年轻时所具有的选择的自由。[168]

法国数学家康尼斯有同样的看法：

[美国体制下的]不断的成果压力缩减了大多数年轻人的“时间单位”。刚入门的人别无选择，只有找一个社会地位牢固的导师（以便

以后为他写相关的推荐信，找一个好工作），然后写一篇专业的学位论文，证明他们有很强的力量；然后还有好多诸如此类的事情。所有这些事情都挤在一个短暂的时间内，他们不可能去学习那些需要几年工夫才能把握的基础东西。我们当然迫切需要技术专家，但那只是促进研究进步的部分条件……在我看来，美国的体制实在打击了真正有创造力的人，那些人在技术上通常是大器晚成的。年轻人在市场上获取职位的方式也成了“诸侯割据”，即有的领域在几所重点大学扎下根来，自我成长，而没有为其他新领域留下成长的空间……结果，只有很少几个学科能得到重视，能持续产生学生，这当然不会为新领域的出现创造良好的条件。169

最近几十年来，商业界已经意识到等级的划分成本太高，于是给年轻人更大的权力和更多的机会。现在有很多年轻的银行家和软件工程师，还有很多人才二十几岁就领导大的项目。偶尔我们也能遇到同样走运的年轻科学家，但很少。许多科学家从博士后的窘迫中脱颖出来，已经是35岁的人了。

高技术公司的领导者们知道，如果想用最优秀的年轻工程师，你就需要年轻的管理者。其他创造性的领域（如音乐）也是如此。我保证某些爵士乐手和老摇滚歌手都欣赏街舞和技术音乐，但音乐公司不会让60岁的老歌星来选择签约的年轻歌手。音乐的创新就踏着这种热烈、活泼的步伐，因为年轻的歌手知道怎样很快与观众和其他歌手沟通，不论在晚会还是电台，而不必求权威的艺术家以他们自己的模式来评判。

有趣的是，量子力学革命完全是孤立的一代物理学家掀起的。他们上一代的许多人都在第一次世界大战中被屠杀了。他们周围几乎没有什么老科学家站出来告诫他们疯了。今天，研究生和博士后为了生存，不得不做能让接近退休的老人明白的事情。这样做科学就像踩着急刹开车。

科学需要叛逆与服从之间的平衡，所以在激进与保守之间总会存在争论。但是现在，科学世界没有平衡，革命比科学历史上的任何时候都处于更凄凉的境地。那样的人简直不为研究大学所容忍。于是，一点儿也不奇怪，即使科学大声呼唤，我们似乎也不能赢得一场革命。

## 第二十章　我们能为科学做什么

我在本书试着解释了为什么物理学的五大难题仍然和30年前的状况一样。为了说明这一点，我不得不集中谈了弦论，但我想重申一下，我的目的不是要丑化它。弦论是一个目标宏大的有威力的思想，为它付出的多数努力都是值得的。如果说它至今尚未成功，主要是因为它的内在缺陷密切关系着它的力量——当然，故事还没完呢，因为弦理论可能成为真理的一部分。真正的问题不在于我们为什么花那么多精力做弦理论，而在于为什么不在其他可能的方法上花同样多的精力。

当年，如果我做量子力学基础，前程可能就毁了；如果做与粒子物理学有关的题目，也许还有希望。当我面临这样的选择时，经济的决定找到了一个科学的理由。显然，过去几十年里，粒子物理学的进步远远超过了量子力学基础的探索。如今，新毕业生的境遇大不相同，局面已经扭转了。过去几十年，粒子物理学没有什么进步，但在量子计算研究的刺激下，基础领域的进步却很大。

现在已经清楚，除非我们认真考察我们对时间、空间和量子的认识基础，否则不可能解决那五大问题；如果我们把像弦理论和圈量子引力那样的老研究纲领当作业已确立的范式，也不可能获得成功。我们需要有勇气、有想象力、有深刻思想并能开辟新方向的年轻科学家。我们当然不能像现在这样打压这样的人，但又该如何识别和支持他们呢？

我必须再次强调，我不认为理论物理学的沉寂应该归咎于哪一个物理学家。我认识的许多弦理论家都是非常优秀的科学家，他们做过很好的工作。我不是要说他们应该做得更好，而只是说我们当中的许多最优秀人物竟然在那么好的思想下面都未能取得成功。

我们面对的是科学世界的社会学现象。我认为科学的规范在很大程度上被第十六章说的那种群体意识玷污了，但那不仅是弦理论群体一个。一方面，制定法则的正是声势浩大的科学群体。在法庭上，好律师会在法律允许的范围内做任何有助于其代理人的事情。我们可以预见，科学领域的领导者也同样会在学术界不成文的规则下做任何能促进其研究项目的事情。如果结果导致一个领域被一系列不成熟的思想垄断了，而那些思想虽然没有什么成果，却凭着空洞的许诺独领风骚，这不能仅仅埋怨科学的领导者，他们也不过是根据他们对科学的认识来行动的。我们可以而且应该将其归咎于所有的科学家，正是他们集体制定了法则来评价同行的声音。

也许很多人会要求我们领域的所有人在接受某个结果之前都要好好检验它。我们可以而且确实把那个任务留给了各个小领域的专家。但我们有责任至少跟踪那些结论和证据。我和许多同事一样，也曾错误地接受了弦理论的一些被广泛认可的东西，尽管它们没有得到科学文献的支持。

于是，我们要问：科学规范的传统约束到底怎么了？我们已经看到，科学结构存在着问题，表现在诸如同行评议和终身教授等制度方面。这对弦理论的垄断起着部分作用，但同样有问题的是常规科学与科学革命的混淆。弦理论开始是一场革命，但现在却被作为常规科学里的另一个研究纲领。

我在前面几章提出，有两种理论物理学家：把握常规科学的工艺师和能洞察假设并提出新问题的预言家。现在已经很清楚了，为了发动科学革命，我们需要更多的预言家。但我们也看到，这些人即使没

有被赶出科学圈，也被边缘化了。他们不再像从前那样被看作主流理论物理学的一部分。如果说我们这一代理论家没有革命，那是因为我们组织科学机构时没有为革命者留下空间，我们多数都听不进那些人的话。

我的结论是，我们必须做两件事情。我们必须认识并反击群体意识的征兆，我们必须向独立思想者敞开大门，确保为革命需要的特殊人才留下空间。这在很大程度上要看我们如何对待下一代。为了维护科学健康，我们应该只凭能力、创造力和独立性来聘用不同的年轻科学家，而不应该看他是否对弦理论或其他研究项目有过什么贡献。甚至还应该有限考虑那些开拓并发展自己研究纲领的人，这样才能使他们能自由探索他们认为最有希望的方法路线。科学管理总是在做选择。为了避免资源过分倾向某个也许是死胡同的特殊方向，大学的物理系应该确保竞争的研究计划和解决问题的不同观点能在教师中得到反映——这不仅是因为我们多数时候不可能预知哪种观点是正确的，也因为在相近领域工作的聪明人之间的友好竞争往往是新思想发现的源泉。

我们还应该提倡一种公开批评、坦诚的态度。要惩罚那些只做肤浅工作而躲避难题的人，奖励那些向长久未决的猜想进攻的人，哪怕要多年以后才能取得进步。在我们统一空间、时间和量子理论的认识的奋斗中，引出了许多基础问题，我们要为那些深入思考这些问题的人留下更多的空间。

我们讨论的许多社会学问题必然与科学家——其实是所有的人——形成部落的倾向有关。为了与这种倾向进行斗争，弦理论家可以弱化弦理论与其他方法之间的界线。他们可以不再根据对这样那样的猜想来划分理论家。弦理论大会应该邀请那些在弦理论的其他路线上工作的人或者批评弦理论的人，集思而广益。研究团队应该找那些在其他方向追寻的博士后、学生和访问学者。还应该鼓励学生学会思考新的解决问题的方法，这样，他们随着学术生涯的进步，才有能力为

自己选择最有前途的方向。

我们物理学家要直面眼前的危机。一个没有预言从而也不管实验的科学理论永远不会失败，但这样的理论也不可能成功，只要科学代表的是来自证据支持的合理论证的知识。一个研究纲领经过几十年还没有发现实验结果的基础或精确的数学形式，我们需要诚实地评估那些坚持这个纲领的智慧。弦理论家需要面对这样的结局：他们错了，而别人对了。

最后，科学的支撑机构可以采取很多步骤以维护科学的健康发展。资助机构和基金会应该让各个层次的科学家探索和开拓各种可行的解决深层次难题的方法。不应该允许一个研究项目在获得可信的科学证明之前成为垄断学科。同时应该鼓励其他方法，这样科学进步才不至于因为在一个错误方向的过分投入而受到阻碍。当出现反对的但是关键的问题时，应该限制给任何单个研究项目的支持比例（例如三分之一）。

有些建议需要重大改革。但是对理论物理学，我们不是说要多少钱。假定某个基金会决定全面资助那些无视量子引力和量子理论主流而探索自己的大胆计划的思想者，那总共也就20几个理论家。资助他们，只需要国家的整个物理学预算的极小部分。但从这些人过去的贡献来判断，也许只有几个人能做出重要的事情而不辜负对本领域的投入。

实际上，从物理学和数学的博士中找那些用自己的方法解决基础问题的人，也就是那些做事情太反传统而很可能永远找不到研究工作的人，只需要更少的钱就能起作用。那些人如巴伯、瓦伦提尼、格罗腾迪克——甚至爱因斯坦。资助他们五年，如果有什么结果，还可以延长到两个或三个五年。

听起来冒险吗？英国皇家学会就有类似的计划。它负责引导那些

在各自领域崭露头角却不太可能得到美国的那种资助的科学家开始自己的研究生涯。

如何选择哪些人值得资助呢？很简单。去问那些这样做过科学的人。为了保险起见，至少要在候选者的领域里找一个对他要做的事情感到兴奋的知名人上。为了完全有把握，还要至少找一个唱反调的教授——他可能认为候选者太可怕，注定要失败。

在一本为大众写的书中讨论科学政策问题似乎很奇怪，但你们（大众）不论个人还是群体都是我们的赞助人。如果说你们花钱支持的科学没有做好，那就需要靠你们驱赶我们去做我们的工作。

所以，最后我要对不同的读者说几句话。

对受过教育的大众：批评。不要相信你听到的多数东西。当一个科学家声称做了什么重要的事情，要他拿出证据来。像你评估投资一样严格评估那些证据。把它当作你要买的房子或者你要送孩子去的学校，仔仔细细地考察它。

对那些决策科学作为的人——如部门领导、人才委员会、基金会官员等：只有你们那个阶层的人才能推荐刚才列举的人。为什么不考虑他们呢？这些建议应该在很多场合展开讨论，如基金会的办公室、国家科学院和世界各地的类似机构。这不仅是理论物理学的问题。如果像物理学那样严格分科的学科都患群体意识综合症，那其他不那么严格的领域又将如何呢？

对我的理论物理学家伙伴：本书讨论的问题我们每个人都有责任。我们成为科学的中坚力量，只是因为我们作为其部分的更大的社会在深切关注着真理。如果弦理论家错了却继续统治我们的领域，后果可能非常严峻——不论对我们个人还是对我们的事业。那就需要我们打开大门，让不同的人走进来，普遍地提高论证的标准。

更坦率地说：如果你是那样的人，当你的科学受到挑战，而你的第一反应是“X怎么想？”或“你怎么能那么说？人人都知道……”那么你恐怕不再是科学家了。你拿那么多钱做工作，意味着你有责任对你和你的同事们相信的每一件事情进行仔细而独立的评价。如果你不能为你的信仰和行为提出恰当的符合实际的理由，如果你要别人来为你考虑（即使他们德高望重），那么你已经违背了你的道德义务，不再属于科学群体的一员了。你的博士头衔给了你坚持自己观点和做出自己判断的资格。但还不仅如此；它还使你有义务批判性地、独立地思考你力所能及的所有事情。

这话有点儿严厉。但对我们做基础问题的非弦理论家来说，还有更严厉的话呢。我们的工作是为了发现错误的假定，提出新的问题，寻找新的答案，发动新的革命。很容易看到弦论可能错在什么地方，但批评弦论不是我们的工作。我们的工作是创立正确的理论。

我对自己会更严厉。我完全相信有读者会反驳我，“既然你那么聪明，怎么不比人家弦理论家做得更好呢？”他们说得很对。因为本书终究不过是一种拖延。我当然希望通过写这本书让感兴趣的人更容易走上那条路。但我的技艺在于理论物理学，我的真正工作是完成爱因斯坦发动的革命，可我没做好。

那么我打算做什么呢？我要利用生活展现给我的好运。首先，我想把我过去的文章“量子与热涨落的关系”找出来好好读一遍。然后，我想切断电话和手机，找一些吉尔博托（Bebel Gil-berto）、艾瑟罗（Esthero）和郎萨克史密斯（Ron Sexsmith）的唱片[170]，把音量调得高高的，擦净黑板，拿一只新粉笔，打开笔记本，拿出我最喜欢的钢笔，坐下来，开始思考。

# 译后记

套一个“经典”的句式说，本书是在物理学30年来没有实质性进步的前所未有的“大不好”形势下产生的。当然也可以说，在物理学新纪元即将来临的“大好”形势下，“个别别有用心的人”跳了出来，攻击多数科学家热情拥护的理论。说实在的，物理学家这些年做了很多事情，发表了很多文章，开拓了很多方向——但在一个无限广袤的“弦景观”面前，他们困惑了。本来是“水光山色与人亲，说不尽无穷好”（李清照句），但物理学家不喜欢“无穷”，只追求惟一。当人们怀着“TheoryofEvery-thing”（TOE）的愿望，迎来的却是“TheoryofNothing”（TON），该是什么感觉啊？—TOE和TON这两个缩写词倒是恰到好处地说明了弦论的历程：“从脚下走出时尚。”在李（LeeSmolin）老师看来，弦论就是一个物理学时尚。一个时代有一个时代的文艺，当然也有一个时代的科学。正如威藤说的，弦论是偶然落在20世纪的21世纪物理学。总之，弦理论大概就代表了我们时代的科学，有人说它是“后现代物理学”，它也的确顺应着后现代的时尚——同一个潮流，多样的选择。它和那些竞争的理论，在我们面前展现了新时代的科学世界的部落文化。本书的意义，也许就在于它体现了这种新的科学文化和精神。

关于超弦理论的异域风光，彭罗斯在大书《通向实在之路》里有一段有趣的比喻：

一个游客来到一座陌生的大城市，要找一幢大楼。那儿没有街名（至少他一个也看不懂），没有地图，在阴沉沉的天空下，他也分不清东西南北。岔路很多，他该左转还是右转？还是走进那条迷人幽深

的小路？街道拐弯很少是直角，路也几乎不是直线；有时走进死胡同，只好退回来，转向别的路。有时会突然发现一条刚才竟然没注意的新路。周围没有人可以打听，他们的话一句也听不懂。不过，游客至少知道他要找的那幢大楼造型精美绝伦，还有漂亮的花园。那也是他去找它的主要原因。他走在看起来很漂亮的路上，有迷人的建筑和美丽的花园，灌木丛生，百花怒放——走近一看，却是塑料的。他面前有很多路，他选择的唯一标准就是那块地方的美感和整体的和谐——风格的和谐或隐藏在城市背后的某种基本的模式。

我们当然是游客，而弦理论家也和游客差不多——老彭的意思是，让他们做导游，还不如我们自己走着瞧呢。现在，李老师充当了导游的角色，不过他是“别有怀抱”的，他要带我们走一条人迹罕至的“迷人幽深的小路”。

李老师以五大问题引领我们的旅行，这些问题是当前隐约看见的物理学的地平线：

问题1：将广义相对论与量子理论结合为一个真正完备的自然理论。

问题2：解决量子力学的基础问题：要么弄清理论所代表的意义，要么创立一个新的有意义的理论。

问题3：确定不同的粒子和力能否统在一个理论并将其解释为一个单独的基本作用。

问题4：自然是如何选择粒子物理学标准模型中的自由常数值的？

问题5：解释暗物质和暗能量。或者，假如它们不存在，那么该如何在大尺度上修正引力理论，为什么修正？更一般地说，为什么宇宙学标准模型的常数（包括暗能量）具有那样的数值？

拿对这些问题的回答给弦理论打分，结果，弦理论似乎只能得1分，就是“潜在地”解决了第三个问题，而对其他问题就不怎么高明了。

当然，我们不能凭问题的回答来决定理论的成败，就像不能凭考卷决定人生。何况问题有不同的提法，不同的导游可以站在不同的石头上讲不同的故事，我们也可以从不同的角度看风景——像《蒙娜丽莎》背后的荒野，左边和右边有着不同的地平线。新世纪到来时，好多科学家提过好多问题——只能问很多，因为谁也不敢把赌注下在具体的某一个。真正“基本”的问题，往往要等回答以后才知道。一个成功的理论可以顺便解决很多“基本”问题；而好的理论从来不是为了同时解决那么多“基本”问题才发展的。《纽约时报》曾假想让科学家沉睡100年，到2100年醒来时向同行打听21世纪的科学。他们竟然一口气问了10个问题。我们可以想象，睡过百年的人睁开眼睛时大概只会问一个问题，可惜我们今天还不知道问哪一个。难怪费曼要说，“我无法确定真正的问题，所以我怀疑没有真正的问题，可我又不敢肯定没有真正的问题。”

接着，李老师回顾物理学的历史——这是一篇浓缩的历史，比很多展开的编年史散发着更为浓烈的物理学的芬芳。以今天的眼光看，物理学发展的历史就是理论统一的历史。不过，从前的统一是自然形成的，而今天的统一是物理学家为自己选择的使命——也是爱因斯坦未竟的事业。回顾历史，是拿历史作镜子。李老师让弦理论暴露在历史的荣光下，是要让它显得黯然失色。

弦理论在短暂的20多年里，经历了两次革命。第一次革命发现弦理论能生出标准模型，第二次革命生出一个M理论的影子。尽管革命了，但不彻底，更远未成功。实际上，“反弦派”从弦理论萌芽的时候就出现了。1985年9月，《科学》杂志的一个编辑请费曼对新的“弦”理论发表意见，费曼写了封有趣的回信：“我不相信它们，但我没有认真研究过它们，也就说不清我为什么不相信它们。”格拉肖认

为弦理论是“中世纪神学的新的翻版”，他还想把弦理论家挡在哈佛的门外（当然失败了）。Rob-ertLaughlin（1998年因为凝聚态物理的贡献获诺贝尔奖）说，“弦理论远非我们更伟大的明天的奇妙的技术希望，而是一个陈旧的信仰系统的悲剧性结果。”这些头顶诺贝尔桂冠的人，为什么不喜欢弦理论呢？大概因为它距离物理现象和实验太远了。20世纪后期的物理学将物理事实变成了数学概念（如规范场等于联络，基本粒子归结为对称），而弦理论就把这些数学结构作为研究对象，得到数学结构的结果。在传统的物理学家看来，这个时尚太形式化；借闻一多先生说初唐诗歌的话，这也算物理学的一种皮肤病。如果说从前物理学是物理学灵魂套上数学外衣，那么弦理论大概就是一个数学灵魂在物理学上空飘浮，总也落不下来。

李老师的立场和那些前辈有些不同。他不是外科医生，而是内科的，甚至精神科的。他是从弦理论阵营里走出来的，他的反对有着更具体的内容。他看到弦理论“腐败”了，今天科学的衰落就是因为弦理论的崛起——所以他的书特别加了一个附标题：“弦理论的崛起与科学的衰落”。当然，李老师不是一个人在战斗。哥伦比亚大学的数学家PeterWoit（也做量子场论）也写了一本书“骂”弦理论，说“好多人在兜售投机的思想”，“弦理论什么结果也没有”，简直“连错都算不上”（notevenwrong，这是他的书名，借用了泡利的名言）。他的副标题正好与李老师的呼应：“弦理论的失败与物理学定律统一的追求”——等于说，弦理论失败了，物理学的统一才可能重新开始。两个不同背景的批判家走到一起了：打倒了弦理论，科学才有希望。仿佛那根弦成了《共产党宣言》里说的套在无产者身上的“锁链”，而李老师们就是无产者和革命者。

李老师说我们正处在可惜的革命时期，而弦理论已经成为“常规”了。从来没有一个理论，经历了那么长的时间、花费了那么多的力量，结果却一无所有。这是李老师对弦理论状况的总评价。当然，仅凭这几点空洞的声讨还不够，因为圈引力也没有令人满意的成绩。

他自己都说，圈量子引力还有很多关键问题没得到解决，就和弦理论中的猜想一样。

所以，李老师对弦理论进行了严厉的技术性批判，主要在三个方面：背景独立、对偶猜想和宇宙学常数。

在圈引力派看来，爱因斯坦广义相对论的核心精神是“背景独立”，即理论不需要一个固定的时空背景。在广义相对论中，时空是与物质相互作用的，物质引起时空弯曲，而时空弯曲表现为物质间的引力。量子论宣扬我们既是演员也是观众，但还有一个舞台，“背景独立”则要让舞台消失在演员的活动中。其实，弦理论家其实也在闹背景独立，但没有拿它作为大旗，他们已经意识到，“为了实现真正的背景独立的形式，似乎还需要更加远离传统的时空观念”。

阿根廷青年Maldacena的猜想是弦理论近10年来最令人激动的思想，它将规范场论作为弦理论的一张全息图，是全息理论的一个数学实现。李老师说它只有“弱形式”（即没有量子效应）才成立，而且怀疑对偶的两个对象是否真的存在。

1998年，宇宙加速膨胀的发现表明存在正的暗能量，但李老师指出弦理论预言暗能量密度（或宇宙学常数）不可能是正的。

2006年12月，圣巴巴拉加州大学的弦理论家JosephPolchinski（昵称Joe，他10年前写的两卷本《弦理论》是流行的弦理论教科书）针对这些问题（连同Woit的那本书），发表了公开“答辩”。

乔老师认为，“背景独立”只是形式问题，不是本质问题。他说李老师“将描述物理的数学预言与被描述的物理混为一谈了。人们常常发现，新物理用的数学语言并不是最恰当的。这一点并不奇怪……在弦理论中，即使语言不是背景独立的，物理也肯定是背景独立的，我们还在继续寻求更恰当的语言。”他认为AdS/CFT对偶可能是问题的

一个解决（当然还不彻底）。在那个猜想下，物理学成了皮影戏——空间的表演都在屏幕上实现。

乔老师从李老师自己写的两篇关于AdS/CFT对偶的论文中发现了概念性的问题，证明李老师的批判完全是误会。他还反过来说李老师所宣扬的圈量子引力其实比眼下的弦理论更加背景相关。

关于宇宙学常数，乔老师认为抓住了李老师的把柄。因为非正的暗能量密度是超对称的预言，而超对称是一定会破缺的。

李老师还一般地批评了弦理论的数学不严密。他寻根溯源，终于考证出文献中从来就没有严格证明过弦理论的有限性，而几乎所有弦理论家却都拿它作为既定的事实。虽然乔老师做过一些计算，但这个问题依然存在。不过乔老师说了，“物理不是数学。物理学家靠计算、物理推理和相互检验，而不靠证明，他们能理解的东西一般说来比能严格证明的东西大得多。”如果要严格以数学的严密来要求，很多物理学恐怕刚一萌芽就凋落了。实际上，不严格的弦理论真的给数学带来了很多东西，也吸引了很多数学家——当然，尽管这是弦理论的功绩，却不能作为它是正确理论的根据。

李老师承认乔老师的“评论是礼貌的，用事实来说话，没有人身攻击，没有误解我们的书（包括Woit的那本）”。不过，2007年4月，他还是“应读者要求”进行了“反答辩”，澄清乔老师“误会”的地方。

关于宇宙学常数问题，李老师同意那是超对称的要求。但他认为，在微扰弦理论中，超对称是清除不掉的。即使在KKTL（以这几个字母打头的四个人发表了一篇文章，他们以非常的工艺，在半经典的情形下构造了具有正宇宙学常数的模型）发表3年以后，“我们仍然不知道在正宇宙学常数背景下是否存在和谐一致的弦理论”。

关于AdS/CFT对偶，李老师又拿数学严格来说话了："我在书中怀疑——现在仍然怀疑——的是，是否证明了猜想的强形式，在AdS5×S5的弦理论和N＝4的规范理论之间建立的等价性。"他说，"证据不代表证明，而这里必须证明两个有明确定义的数学对象之间的等价性……我们没有严格的非微扰的在渐近AdS5×S5背景下的弦理论的定义，也没有严格的非微扰的4维N＝4超对称杨一米尔斯规范理论的定义。没有定义，我们甚至不能肯定是否能很好定义与那些名词相应的数学结构。"（从这段话我们也可以看出，弦理论的确在研究数学结构。）

关于背景独立问题，李老师承认强形式的Maldecena猜想（如果正确的话）可能提供"非常有限的弱形式的背景独立"。他还纠正了乔老师"在弦理论中"那句话，认为应该更准确地这样说：

有些弦理论家相信，他们具体研究的不同形式的微扰弦理论及其对偶接近某个更深刻的背景独立的形式。这缺失的背景独立形式并不仅仅是不同的语言，而有可能表达了确定理论的原理和定律，迄今所研究的一切都将作为近似从它们推导出来。尽管弦理论家们相信这一点，但对这个猜想的背景独立形式的弦理论，他们只提出过寥寥几个具体的设想，而且没有得到广泛的支持。

两位老师的话正好体现了他们对弦理论的不同态度。在李老师看来，乔老师们显然过高估计了弦理论的成绩。

2007年5月20日，乔老师发表了给李老师的信，更明确地反对了他的反对。他继续批评李老师对弦理论的认识太落伍了。李老师在书中的一句话也可能伤了弦理论家们的心："…一大群专家在尽力挽救一个他们珍爱的然而却面对着矛盾数据的理论。"（第十章）乔老师好像有点儿生气了："这样的例子在你的书中俯拾皆是：你写的是你自己相信的事情，你却忽略了事实……明明是成功的地方，你却说危机；本来没有的问题，你却讲伦理。"

我们“津津有味”转述两位的争论，一方面是为了补充作者在书中没有说完的话；另一方面是为了“兼听”不同的声音。对台戏总比独角戏有意思。虽然乔老师们没有写一本书，但带着他们的问题去跟李老师走，看周围的风景就不会是一个色调了。

弦理论的争论还在以多种形式继续着。PaulC.W.Davis（《上帝与新物理学》等畅销书的作者，如今把兴趣转向了天体生物学）在《纽约时报》（2007年11月24日）发表文章说，物理学家们的“时尚变了”，变得像宗教一样，“拿信仰做基础”，相信存在“一个巨大的看不见的宇宙的集合”（即所谓的“多重宇宙”）。Woit借题发挥说，物理学家真正的改变在于失去了弦理论的兴趣，因为它的支持者们只有借多重宇宙的假说来解释他们为什么不能做出任何预言。数学的“物理学定律”以可检验的方式描写世界，这不是信仰，而是事实。

最近，著名的“第三文化”的论坛“边缘”（www.edge.org）提出了新的年度问题：

当思想改变你的思想，那是哲学；

当上帝改变你的思想，那是信仰；

当事实改变你的思想，那是科学。

那什么改变了你的思想呢？

落实到弦理论，它改变我们什么了呢？它似乎没谈多少“事实”，那么它是科学，是哲学，还是信仰？本来很简单的问题，在一定的框架下追问，就复杂而且严肃起来了。160多个（还不断有人加入）不同领域的学者参加了问答，畅谈了自己的感受。其中一个物理学家（达特茅斯学院的MarceloGleiser）的话，说得很“实在”：

我从小就被灌输统一的思想。它首先是来自宗教……十几岁时，

我开始对科学感兴趣，开始怀疑无处不在的上帝，怀疑大洪水、戒律和瘟疫的故事，转向了物理学，把爱因斯坦和他的科学当偶像……

做研究时，我毅然决定做一名理论物理学家，做粒子物理学和宇宙学，原因很简单：它是极大和极小的两个世界的结合，最有可能发现自然的统一理论……

我写过几十篇与统一有关的文章，连博士论文也是那个题目。我曾为那些思想的现代方法着迷了：超对称、超弦、隐藏的额外维的空间……可是，几年前，也许因为我更深刻认识了形成科学思想的历史和文化过程，事情突然变了。我开始怀疑统一，觉得它不过是实在的一神论在科学的翻版，是在方程里寻找神的存在……二十多年过去了，所有的努力都失败了。粒子加速器没有，冷暗物质探测器也没有，没找到磁单极，没看到质子衰变，过去几十年预言的所有统一的迹象，都没有……

他还说了句爱因斯坦式的格言：“自然不欣赏我们的神话。”我们真是在写新的神话吗？如果我们在用数学写神话，那么似乎应该听听一个老数学家的最新感悟：“归根结底，数学只是我们的，而不是宇宙的。”

同样背离初衷的还有JoneBaez，他曾做过多年的圈引力，但现在对弦和圈都失去了信心，最后决定不做量子引力了。“这是非常痛苦的决定，因为量子引力曾是我追寻了几十年的圣杯。”不过，现在他感觉彻底解放了。

李老师也参加了问答，但没有说圈和弦，而是谈自己对时间的认识的改变。他的结论是，时间也许不是幻觉，而是实在。他认为，时间的本质关联着数学真理的本质，也关联着是否存在没有时间的自然定律。当然，这些问题在深层意义上也是未来的理论必须回答的。

不久之前，弦理论的元老施瓦兹在汤川秀树一朝永振一郎百年诞

辰的纪念会上，总结了弦理论的形势和任务。虽然弦理论已经取得了许多进步，“但即使继续以当前的步伐快速向前，我想这门学科到了纪念两先生二百周年的时候也不会完成。”这样说来，威藤的话大概还要改两个数字：弦理论是22世纪的理论落到21世纪的部分。

听着这些热烈的争论和深沉的反思，我们不禁想起狄更斯的名句（《双城记》）：

那是最美好的时代，那是最丑恶的时代；那是智慧的年月，那是愚昧的年月；那是信仰的时期，那是怀疑的时期；那是光明的季节，那是黑暗的季节；那是希望的春天，那是失望的冬天……

这就是我们的“后现代物理学”的时代吗？

本书的另一个主题（大约四分之一的篇幅）是关于科学的伦理或科学的社会学。乔老师在评论中说，李老师“将弦理论的兴衰看作一出道德剧”。李老师自己也将“什么是科学，科学如何运作”的话题作为本书的关键（第十七章）。他将弦理论作为科学群体的伦理活动的一个案例来研究，这一点，不管弦理论和圈引力的命运如何，也许都将是未来科学史感兴趣的课题，而本书提供了极好的案例。

李老师的科学伦理观，源于他本人的经历。他讲了几个故事。一个是在做研究生时从超引力逃出来（第六章），他的觉悟是（一段金玉良言）：

当我回顾30年来我所熟悉的那些人们的科学生涯时，越发感到科学生涯的抉择依赖于人的个性。有些人乐于跨越下一步，把一切都献给它，从而为飞速发展的领域做出重要贡献。另一些人可没那么急躁。有些人容易犯糊涂，所以做什么都要反复思量，这要费很长的时间。你大概以为我们比这些人高明，可别忘了爱因斯坦也是属于他们的。根据我的经验，真正令人震撼的新思想方法往往来自这样的人

群。还有一些人——我属于这第三类——只顾走自己的路，他们特立独行，只是因为不愿意像有的人那样为了站在赢家的一边而加入某个领域。所以，当我与别人的作为相左时，也不再感到烦恼，因为我发现一个人的性情几乎完全决定了他做什么样的科学。幸运的是，科学需要来自不同类型的人物的贡献。我逐渐认识到，那些能把科学做好的人是因为他们选择了适合自己的问题。

第二个故事是他在伯克利和“无政府主义”哲学家费耶阿本德的谈话。费老的经历很有趣，他告诫他们，“就做你想做的，不要管别的事情。在我的经历中，从来没有花过五分钟做我不想做的事情。”（第十七章）费老的卓尔不群正好在他心头引起了共鸣。他从费老那儿学到了“反对方法”。作为严肃的物理学家，他当然不是要反对科学的方法，而是借一面“无政府”的大旗来反对弦理论的“垄断”。李老师为年轻人说了很多好话。他批评弦理论聚集了那么多人，垄断了那么多资源，却伤害了年轻人，扼杀了新思想和新发现。他把物理学家分成两类，预言家与手艺人（seersviruscraftspeople）—思想家与匠人。当然不能说做常规科学的人就没有思想，李老师强调的是与常规“格格不入”的思想，能引领未来的思想（所以叫“预言”）。他提出几个主流外的“预言家”的典范——如躲到乡间独自思考的巴伯，身为教授而志在江湖的芬克尔斯坦，在意大利游荡的霍金的小师弟瓦伦提尼……所有这些人，都有一个共同的基本的经历，那就是“目前健在的最能干、最富想象的数学家”格罗藤迪克（Alexader Grothendieck）说的：

他们必须重新发现自身的那种与生俱来的（和我一样的）能力：忍受孤独。

李老师还特别以“修正的相对论”（DSR）作为新理论的样本（他从它引出了一个有趣的“虹宇宙”的图景）。科学总是需要不同的声音，特别在“万马齐喑”的时候。最近，密执安大学ScottE.Page写了一本书，谈思想的多样性能产生更具创造性的团队、公司、学派和社会

（“The Difference：How the Power of Diver-sity Creates Better Groups, Firms, Schools and Societies”，Princeton University Press，2007）。其实，很多人都知道这一点。但我们难免和李老师一样奇怪：科学的管理者们为什么不那么做呢？

最近，英国高等教育基金会（HEFCE）决定从2009年起用一种“度量体系”来取代同行评议制度，即以研究成果（论文）的引用数作为评判标准。这当然是更不利于李老师同情的那些叛逆者。

李老师如今在主持“圆周（Perimeter）”——即“圆周理论物理研究所”，它的标志和缩写是PI，即π。PI座落在加拿大多伦多外的沃特卢，发起者和赞助人是RIM公司（Blackberry的制造商）的Michael Lazaridis，他原来的想法是要做一个普林斯顿式的高等研究机构。2001年9月开张时，只有三个固定人员：李老师和Robert Myers, FotiniMarkopoulou。PI现在是量子引力和量子信息理论的热土。2007年，加拿大自然科学与工程研究理事会（NSERC）对PI的评价是，“自1999年成立以来，PI已成为加拿大在新兴量子物理领域的龙头和科学教育和服务的典范。”这是一个新兴的机构，从老牌大学的官僚作风里走出来，自然感觉是“解放了”。

但是解放者的声音，压迫者是听不进的。《淮南子?人间训》讲过一个故事（后来冯梦龙把它编在《智囊》的第一篇）：孔子的马偷吃了农夫的庄稼，农夫就把马关起来了。子贡跑去求情，低声下气，人家还不理他。夫子批评他说：“夫以人之所不能听说人，譬以大牢享野兽，以《九韶》乐飞鸟也……”李老师的话，对那些科研管理人员和决策者来说，大概就是鸟儿耳朵里的韶乐。其实，管理者们也有自己的逻辑和道理，而且也在像做弦理论一样做“科学的科学”。不过，他们学得最好的还是欧几里得几何，特别擅长以公理的方法来抹平事情。他们似乎发现，最科学的“科学管理”，就是尽可能地“统一标准”，当一切的不公平都以法规的形式定下来后，就没有不公平了。李老师们在博士后经历的磨难，我们的同学从幼儿园就开始经历了。

本文拖拉着要结束时，看到一个消息，或许印证了李老师在本书前言里的判断：我们今天并不比30年前懂得更多。最近，一个国际研究小组用超级计算机验证了标准模型过去的预言，但没有发现任何新的东西。爱丁堡大学RichardKenway教授说：“尽管标准模型取得了巨大成功，但因为必要的计算尚不够精确，还有一两个角落没得到确切的实验证明。我们照亮了其中一点，但令人泄气的是，什么也没看见。”

译者

2008年4月9日于重庆

[1]Mark Wise，“Modiacations to the Properties of the Higgs Boson，”Seminar talk, Mar.23，2006.Available at http://streamer.perimeterinstitute.ca：81/-mediasite/.

[2]动力三角化是荷兰乌特列兹大学的Renate Loll，丹麦哥本哈根大学Jan AmbjØrn和波兰亚格隆尼大学Jerzy Jurkiewicz提出的时空几何，主要思想是将时空剖分为微小的三角形结构。斯莫林似乎很看好这个模型，——译者

[3]Brian Greene.The Fabrie of the Cosmos：Space, Time, and the Texture of Reality（New York：Alfred A.Knopf，2005），p.376.（此书可以认为是格林为《宇宙的琴弦》写的续篇，中译本也将纳入“第一推动”丛书。）

[4]Gerardt Hooft, In Search of the Ultimate Building Blocks（Cambridge：Cambndge University Press，1996），p.163.

[5]引自2005年10月10日《新科学家》杂志：“诺贝尔奖得主承认弦理论陷入困境。”文章引发了一定的争议，所以格罗斯在第23期耶路撒冷冬季理论物理学校的开幕词上澄清了自己的言论（全文见www.as.huji.ac.il/schools/phys23/media.shtml）：我的话的真正意思是，我们还不知道弦理论是什么，不知道它是不是最后的理论，或者它还缺少什么东西，而我们必须等着面对概念的深刻改变……特别是空间和时间本性的改变。但这并不是说我们应该停止做弦理论了——它失败了，完蛋了——不是的，我们正处在一个神奇的时期。

[6]J.Polchinski, Talk given at the 26th SLAC Summer Institute on Particle Physics，1998，hep-th/9812104.

[7]http://motls.blogspot.com/2005/09/why-no-new-einstein-ii.html.

[8]Lisa Randall，“Designing Words，”in Intelligent Thought：Science Versus the Intelli-gent Design Movement, ed.John Brockman（New York：Vintage，2006）.

[9]John Stachel，“How Did Einstein Discover Relativity?”http://www.aip.org/-history/einstein/essay-einstem-relativity.htm.

[10]读者应该注意我这儿讲的已经大大简化了。还有些关键实验，涉及光通过流水或光行差效应（地球和恒星的相对运动对星光观测的影响）。爱因斯坦也不是唯一认识到正确答案应该融合相对性原理的人，伟大的法国数学家和物理学家庞加莱（Henri Poincarè）也认识到了这一点。

[11]我应该承认，诺德斯特罗姆不是这样解决问题的，但他可能这样做。这是后来拥护多维空间的人所用的方法，比诺德斯特罗姆更进了一步。

[12]有一点需要注意的是，这个原理只适用于小时间间隔内在小空间区域的观察。如果下落距离很远，引力场强度发生了变化，引力与加速度就可以区分开了。

[13]这里，专家也许喜欢更准确的惯性概念，但那可能令外行的读者感到迷糊。

[14]当然，正如我们已经看到的，暗能量和暗物质的预言例外。

[15]引自Huhert F.M Goenner, On the History of United Field Theories（1914～1933），p.30.http://relativity livingre views.org/Articles/lrr-2004-2/index.html（2004）.

[16]同上，Hubert F.M.Goenner, pp.38～39.

[17]同上，Hubert F.M.Goenner, p.39.

[18]同上，Hubert F.M.Goenner, p.35.

[19]引自AbrahamPais, SubtleistheLord（NewYork：OxfordUniv.Press，1982），p.330.

[20]同上，p.332。

[21]同上。

[22]同上，p 334。

[23]这说的是爱因斯坦从纳粹德国出来。——译者

[24]CERN是法文Conseil Europèen pour la Recherche Nuclèaire（欧洲核子研究理事会）的缩写，成立于1954年，是世界最大的粒子物理研究中心，由20个成员国提供资金，树立了国际科学合作的典范——译者

[25]有兴趣进一步学习的读者，可以在我1997年《宇宙的一生》（The Life of the Cosmos, New York：Oxford University Press）第四章阅读有关规范对称性的东西。

[26]尽管不需要，但也许有读者想知道规范原理是如何起作用的。思想的关键是这样的：决定对称的操作通常需要作用于整个系统。为了说明一个物体在旋转下是对称的，我们需要旋转整个物体。你不可能只转一个球的某个部分。但在特殊情形下，对称性也能只作用于系统的某个部分。这种对称性叫局域对称性。这似乎与直觉矛盾，怎么会这样呢？原来（这一点离开数学很难解释），当系统的不同部分以某种力发生相互作用时，它就能起作用那些力就是规范力。

[27]这段历史同样比我概括的要复杂得多。杨-米尔斯理论首先是在20世纪20年代的高

维统一理论的背景下发现的，但被人遗忘了。20世纪50年代，杨振宁和米尔斯（Robert Mills）才重新发现了它。

[28]Y Nomura and B.Tweedie, hep-ph/0504246

[29]弗拉姆普顿的电子邮件（经允许引用）。

[30]A.Einstein，“Approximate Integration of the Field Equations of Gravitation，”Sit-zungsberichte der Preussische Akadamie der Wissenschaften（Berlin，1916）688～696.关于量子引力的早期历史，见John Stachel（引言和第五部分评论），Conceptual Foundations of Quantum Field Theory, ed.Tian Yu Cao（Cambridge：Cambridge University Press，1999）

[31]W.Heisenberg and W.Pauli，“Zur Quantendynamik der Wellenfelder，”Zeit.fü r Physik，56：1～61（1929），p.3.

[32]M.P.Bronstein，“Quantization of gravitational waves，”Zh.Eksp.Teor.Fix.6（1936）195.关于Bronstein的详细情况，见Stachel in Conceptual Foundations, and als。C.Gorelik，“Matvei Bronstein and Quantum Gravity：70th Anniversary of the Unsolved Prob-lem，”Physics-Uspekhi，48：10（2005）.

[33]Richard P.Feynman, What Do You Care What Other People Think?（New York：W.W.Norton，1988），p.91.

[34]其实这是引力束缚系统（诸如恒星和星系）的一般性质。那些系统在输入能量时会发生冷却。引力系统与非引力系统的这个根本区别正是很多统一物理学的尝试所遭遇的巨大绊脚石。

[35]G.Veneziano，“Construction of a Crossing-Yymmetric Regge-Behaved Amplitude for Linearly Rising Regge Trajectories，”Nuovo Cim.，57 A：190～197（1968）.

[36]http://www.edge.org/3rd_culture/susskind03/susskind_index.html.

[37]P.Ramond，“Dual theory for free fermions，”Phys.Rev.D，3（10）：2415～2418（1971）.

[38]另一篇特别重要的论文是“Quantum Dynamics of a Massless Relativistic String，”by P.Goddard, J.Goldstone, C.Rebbi, and C.Thom, Nu.cl.Phys.，56：109～135（1973）.

[39]J.Scherk and J.H.Schwarz，“Dual Models for Non-Hadrons，”Nuc.Phys.B，

81（1）：118~144（1974）.

[40]T.Yoneya，“Connection of Dual Models to Electrodynamics and Gravidynamics，”Prog.Theor.Phys..51（6）：1907~1920（1974）.

[41]一个理论在数学上和谐，指的是它不会给出自相矛盾的结果。一个相关的要求是，理论描述的所有物理学量都应该是有限的。

[42]J.H.Schwarz, interviewed by Sara Lippincott, July 21 and 26，2000，http://-oralhistories.library.caltech.edu/116/01/Schwarz_OHO.pdf.

[43]M.B.Green and J.H.Schwarz，“Anomaly Cancellations in Supersymmetric D＝10 Gauge Theory and Superstring Theory，”Phys.Lett.B，149（1~3）：117~122（1984）.

[44]施瓦兹访谈。

[45]Thomas S.Kuhn, The Structure of Scientiac Revolutions（Chicago：Univ.of Chicago Press，1962）.

[46]S.Mandelstam，“The N-loop String Amplitude—Explicit Formulas, Finitenness and Absence of Ambiguities，”Phys.Lett.B，277（1~2）：82~88（1992）.

[47]P.Candelas et al.，“Vacuum Conagurations for Superstrings，”Nuc.Phys.B，258（1）：46~74（1985）.

[48]A.Strominger，“Superstrings with Torsion，”Nuclear Physics B，274（2）：253~284（1986）.

[49]In P.C.W.Davies and Julian Brown, eds.，Superstrings：A Theory of Everything（Cambridge Univ.Press，1988），pp.194~195.

[50]Sheldon L.Glashow and Ben Bova, Interactions：A Journey Through the Mind of a Particle Physicist（New York：Warner Books，1988），p.25.

[51]L.Smolin，“Did the Universe Evolve，”Class.Quantum Grav.，9（1）：173~191（1992）.

[52]更通常的说法是，T代表targetspace（“靶空间”）。——译者

[53]E.Witten，“String Theory Dynamics in Various Dimensions，”hepth/9503124；Nucl.Phys.B，443：85~126（1995）.

[54]C.M.Hull and P.K.Townsend，“Unity of Superstring Dualities，”hepth/9410167；Nucl.Phys.B，438：109～137（1994）.

[55]私下交流。

[56]J.Polchinski，“DirichletBranes andRamond-RamondCharges，”Phys.Rev.Lett.，75（26）：4724～4727（1995）.

[57]J.Maldacena，“The Large N Limit of Superconformal Field Theories and Supergravi-ty，”hep-th/9711200；Adv.Theor.Math.Phys.，2：231～252（1998）；Int.J.Theor.Phys.，38：1113～1133（1999）.

[58]A.M.Polyakov，“A Few Projects in String Theory，”hep-th/9304146.

[59]B.de Wit, J.Hoppe, and H.Nicolai，“On the Quantum-Mechanics of Supermem-branes，”Nuc.Phys.B，05（4）：545～581（1988）.

[60]T.Banks, W.Fischler, S.Shenker, and L.Susskind，“M-Theory as a Matrix Model：A Conjecture，”Phys.Rev.D，55（8）：5112～5128（1997）.

[61]超新星观测是SaulPerlmutter和他在劳伦斯伯克利实验室的合作者以及Robert Kirschner和他的“大红移超新星搜寻”小组完成的。

[62]E.Witten，“Quantum Gravityinde Sitter Space，”hep-th/0106109.威藤接着说，“考虑到经典的止步定理，最后这个陈述并不十分令人惊奇。因为，从稳定模的通常问题来看，鉴于德西特空间不会经典地产生，也很难在量子水平上以可靠的方式得到它”（所谓经典的止步定理，指SidneyColemanandJeffreyMandula在1967年提出的no-go定理，说一个和谐的4维量子场论的对称群只能是内禀对称群与Poincare群的直积，而“不可能以任何非平凡的方式混合时空和内在对称性”。它在超对称性的推广即Haag-Lopuszanski-Sohmus定理。——译者）

[63]S.Kachru, R.Kallosh, A.Linde, andS.Trivedi，“De Sitter Vacuain String The-ory，”hep-th/0301240.（这篇论文被广泛引用，简称KKLT。——译者）

[64]例如，参见T.Hertog, G.T.Horowitz, and K.Maeda，“Negative Energy Density in Calabi-Yau Compatifications，”hep-th/0304199，Jour.High Energy Pgys.，0305：60（2003）.

[65]L.Susskind，“The Anthropic Landscape of String Theory，”hep-th/0302219.

[66]原来，在这儿用了一个错误的原理：假定我们观测O并考虑两个可能的解释。在给定解释A的情况下，0的概率很低，但在给定解释B的情况下，O的概率很高。由此很容

易得出结论说B的概率比A的概率高，但没有一个逻辑或概率原理允许这样的推论。

[67]S.Weinberg，“Anthropic Bound on the Cosmological Constant，”Phys.Rev.Lett.，59（22）：2607～2610（1987）.

[68]L.Smolin，“Did the Universe Evolve?”Class.Quantum Grav.，9（1）：173～191（1992）.

[69]Wemberg，“Living in the Multiverse，”hep-th/0511037

[70]引自SeedMagazine最近对人存原理与弦理论泛滥之间的关系的考察，见http://www.seed magazine.com/news/2005/12/surveying_the_landscape.

[71]E.J.Copeland, R.C.Myers, and J.Polchinski，“Cosmic F-and D-Strings，”Jour.High Energy Phys.，Art.no.013，June 2004.

[72]M.Sazhi n et al.，“C SL-1：Ch ance Pro jection Effect or Serendipitous Discovery of a Gravitational Lens Induced by a Cosmic String?”Mon.Not.R.Astron.Soc.，343：353359（2003）.

[73]N.Arkani-Hamed, G.Dvali, and S.Dimopoulos，“The Hierarchy Problem and New Dimensions at a Millimeter，”Phys.Lett.B.，429：263～272（1998）.

[74]L.Randal l and R Sundrum，“An Alternative to C.ompactiacation，”hep-th/9906064；Phys.Rev.Lett.，83：4690～4693（1999）.

[75]从技术上讲，超对称意味着在时空几何中存在一个类时或类光的Killing场，这意味着存在时间对称性，因为（专业地说）超对称代数在Hamilton量上是闭的。换句话说，超对称性需要一个Killing旋量，隐含着零或类时Killing向量。

[76]E.D’Hoker and D.H.Phong, Phys.Lett.B，529：241～255（2002）；hep-th/0110247.

[77]轴子（axion）原是某些大统一理论要求的（据说是以洗涤粉命名的）小质量中性粒子，质量小于10-5eV。大爆炸也许为宇宙留下了非常多的轴子，从而可能使时空平直，于是成为暗物质的候选者。根据2007年7月的意大利PVLAS（专门探测暗物质的实验装置）实验结果，还没有发现轴子。——译者

[78]D.Friedan，“A Tentative Theory of Large Distance Physics，”hep-th/0204131.

[79]D.Karabali, C.Kim, and V.P.Nair, Phys.Lett.B.434：103～109（1998）；hep-th/9804132；R.G.Leigh, D.Minic, and A.Yelnikov, hep-th/0604060.对3＋1维的应

用，见L.Freidel, hep-th/0604185.

[80]最近，这些新技术也成功用到了现实世界的3个空间维的QCD。

[81]彭罗斯在《通向实在之路》（TheRoadtoReality，2005——中译本即将由湖南科学技术出版社出版）中指出，额外维卷缩成的多数紧化空间都将很快坍缩成奇点。为说明这一点，他将他本人与霍金发现的定理（即广义相对论预言了宇宙学解的奇点）用于这些弦理论的时空背景。据我所知，他的论证是正确的。它们仅在经典的近似水平成立，但我们也只能用这种近似方法来研究弦理论中时空背景的时间演化。于是，彭罗斯的结果与令弦理论家们相信存在弦理论景观的论证是同样可靠的。

[82]引自Amanda Gefter，“Is String Theory in Trouble”New Scientist, Dec.17.2005.

[83]事情常常是这样的：当其他实验家重复实验时，并不能验证那些令人惊奇的实验结果。这并不意味着有谁不诚实。可能事物的实验几乎总是不可复制的，很难区分噪声与有用的信号。通常需要不同的人用很多年、经过很多努力，才可能在新实验中认识和清除所有的误差来源。

[84]用R来表示，宇宙学常数等于1/R2。

[85]K.Land and J.Magueijo，“Examination of evidence for a Preferred Axis in the Cosmic Radiation Anisotropy，”Phys.Rev.Lett，95：071301（2005）.

[86]K.Land and J.Magueijo，“Examination of Evidence for a Preferred Axis in the Cosmic Radiation Anisotropy，”Phys.Rev.Lett，95：071301（2005）.

[87]M.Milgrom.“AModification of the Newtonian Dynamics as a Possible Alternative to the Hidden Mass Hypothesis，”Astrophys.Jour，270（2）：365～389（1983）.

[88]关于MOND的更多信息和支持它的证据连同相关文献，见www.astro umd.edu/～ssm/mord/.

[89]J.D.Anderson et al，“Study of the Anomalous Acceleration of Pioneer 10 and 11，”gr-qc/0104064.

[90]实际上（有趣），现在的“米”是用光速来定义的，而光速就定义为那个常数c显然其中正潜藏着这里讨论的问题——译者

[91]M.T.Murphy et al.，“Further Evidence for a Variable Fine Structure Constant from Keck/HIRES QSO Absorption Spectra，”Mon.Not.Roy Ast.Soc.345：609～638（2003）.

[92]见E.Peik et al.，“Limit on the Present Temporal Variation of the Fine Structure constant，”Phys.Rev.Lett，93（17）：170801（2004）和R.Srianand et al.，“Limits on the Time Variation of the Electromagnetic Fine Structure Constant in the Low Energy Limit from Absorption Lines in the Spectra of Distant Quasars，”Phys.Rev.Lett.，92（12）：121302（2004）.

[93]K.Greisen，“End to the Cosmic-Ray Spectrum?”Phys.Rev.Lett.，16（17）：748～750（1966）and G.T.Zatsepin and V.A.Kuzmin，“Upper Limit of the Spectrum of Cosmic Rays，”JETP Letters，4：78～80（1966）.

[94]S.Coleman and S.L.Glashow，“Cosmic Ray and Neutrino Tests of Special Relativity，”Phys.Rev.B，405：249～252（1997）；Coleman and Glashow，“Evading the GZK Cosmic-Ray Cutoff，”hep-ph/9808466.

[95]G.Amelino-Camelia，“Testable Scenario for Relativity with Minimum-Length，”hep-th/0012238.

[96]Joao Magueijo, Faster Than the Speed of Light：The Story of a Scientiac Speculation，（New York：Perseus Books，2003）.（有中译本，赵文译，湖南科学技术出版社，2005——译者）

[97]Vladimir Fock, The Theory of Space, Time, and Gravitation（London：Pergamon Press，1959）.

[98]平钦（1937～）是美国小说家，《引力之虹》是他1973年的作品。从DSR导出的时空度规不是唯一的，而是与能量有关的一组度规，犹如彩虹是不同能量的一组光线，所以叫“虹度规”。而在广义相对论中，度规就是引力，所以理论得了这个美丽动听的名字。参看作者的同名文章：JoāoMagueijoandLeeSmolin，“Gravitys Rainbow，”gr-qc/0305055.——译者

[99]L.Friedel, J.Kowalski-Glikman, and L.smolin，“2＋1 Gravity and Doubly Special Relativity，”Phys.Rev.D.，69：044001（2004）.

[100]E.Livine and L.Friedel，“Ponzano-Regge Model Revisited　III：Feynman Diagrams and Effective Field Theory，”hep-th/0502106；Class.Quant.Grav.，23：2021～2062（2006）.

[101]Florian Girelli and Etera R.Livine，“Physics of Deformed Special Relativity，”gr-qc/0412079.

[102]A.Ashtekar，“New Variasbles for Classical and Quantum

Gravity，”Phys.Rev.Lett，57（18）：2244～2247（1986）.

[103]http://online.kitp.ucsb edu/online/kitp 25/witten/oh/10.html.

[104]这并不总是普遍相信的；坚持因果性作用的主要人物有RogerPenrose, Rafael Sorkin, FayDowker和FotiniMarkopoulou。

[105]见，例如RLoll, J.AmbjØrn, andJ.Jurkiewicz，“The Universe from Scratch，”hep-th/0509010.

[106]富勒是20世纪罕见的通才，是哲学家、思想家、幻想家、发明家、建筑师、工程师、数学家、诗人、宇宙学家等。当然，他最为大众熟悉的成就还是由基本三角形构成的多面体（或球形）的穹隆结构（C60具有类似穹隆的结构，因而叫富勒烯）。——译者

[107]见，例如Alain Connes, Noncommutative Geometry（San Diego：Academic Press，1994）.

[108]O.Dreyer，“Background-Independent Quantum Field Theory and the Cosmological Constant Problem，”he-p-th/0409048

[109]见，例如，C.Rovelli，“Graviton Propagator from Background-Independent Quan-tum Gravity，”gr-qc/0508124.

[110]s.Hofmann and O.Winkler，“The Spectrum of Fluctuations in Singularity-free Inflationary Quantum Cosmology，”astro-ph/0411124.

[111]F.Markopoulou，“Towards gravity from quantum，”hep-hp/0604120.

[112]s.O.Bilson-Thompson，“A Topological Model of Composite Preons，”hep-ph/0503213.

[113]S.O.Bilson-Thompson, F.Markopoulou, and L.Smolin，“Quantum Gravitv and the Standard Model，”hep-ht/0603022.

[114]讲座的录音可以听http://www.Perimeterinstitute-.ca/activities/scientiac/cws/evolving_laws/.

[115]www.cosmicvanance.com/2005/11/18/a-particle-physicists-perspective.

[116]一个不知名者的帖子，见2004年12月9日http://groups.goodle.com/group/sci.physics.strings/.

[117]2005年1月20日GuardianUnlimited

[118]针对我对弦理论某些结果的质问，我收到过三个回答，回信人都提到“强”理论群体。例如，“虽然微扰的有限性（或马尔德希纳猜想、s对偶）也许尚未得到证明，但没有一个强理论群体的人会相信它是错的”说一次可能是偶然，说三次就是经典的弗洛伊德式的口误了。人总是渴望成为最强群体的一员，这一点能从弦理论的社会学看出多少呢？

[119]s.Kachru, R.Kallosh, A，Linde, and S.Trivedi，“De Sitter Vacua in String Theory，”hep-ht/0301240.

[120]http://groups.goodle.com/group/sci, physics.strings/，April6，2004.

[121]L.Smolin，“DidtheUniverseEvolve?”Class, Quant.Gtav.，9：173～191（1992）.

[122]八元数是对四元数的推广，是JohnT.Graves（四元数发明者WilliamHamil-ton的朋友）在1843年发现的；1845年，ArthurCayley也独立发表了一篇关于它的论文八元数既不满足交换集，也不满足结合律，因而被长期冷落，不过如今在相对论、弦理论和量子逻辑等领域发挥了作用。——译者

[123]私下交流。

[124]MichaelDuff, PhysicsWorld, Dec.2005.（其实，Duff的这段文字，除了最后一句话，早就出现了2000年12月的《物理世界》，评说的对象正是斯莫林自己的ThreeRoadstoQuantumGravity.MichaelDuff当时是密歇根大学OskarKlein物理学教授，现在是伦敦帝国学院AbdusSalam理论物理学教授。他在文章里说斯莫林的思想“总是在主流之外”，“所以读者应该明白，斯莫林的观点是非常怪异的，绝不能代表当前的思想。”——译者）

[125]www.damtp.can.ac.uk/user/gr/public/qg_ss.html.

[126]不过，我很高兴地报告大家，在网上也不难找到没有歪曲或夸张的弦理论介绍。下面是几个例子：http://tena4.vub.ac.be/beyondstnngtheory/indes.html；http://www.sukdon.com/jpi-erre/strigs/；http://en, wikipediaorg/wik/M-theory

[127]S.Mandelstam，“The N-loop String Amplitude-Explicit Formulas, Finiteness and Absence of Ambiguities，”Phys.Lett.B，277（1～2）：82～88（1992）.

[128]下面是几个例子：J Barbon, hep-ht/0404188，Eur Phys.J.，C33：S67-S74（2004）；S.Foerste, hep-th/0110055，Fortsch.Phys.，50：221～

403（2002）；S.R.Giddings, hep-th/0501080；andI.AntoniadisandG.Ovarlez, hep-th/9906108还有一篇难得的评论，谨慎而正确（就当时而言）地讨论了有限性问题，即L.Al-varez-GaumeandM.AVazquez-mozo, hep-th/9212006

[129]这篇文章是Andrei Marshakov写的（Phys.Usp，45：915～954（2002），hep-th/0212114）很抱歉下面一段话的术语太多了，不过也许有的读者会看出关键思想的：遗憾的是，自以为在现有弦论模型中最成功的十维超弦，一般说来只有在树和单圈水平上才能严格定义。从两圈联络到散射振幅，做扰弦理论中的所有表达式都没有真正定义。原因来自超列称几何或复结构模在超对称伙伴的积分这与玻色弦的情形不同，那儿的积分度量由Belavin-Knizhnik定理决定，积分度量在超模（更严格说，超复结构的奇模）上的定义仍然是一个悬而未决的问题[88，22]，黎曼曲面复结构的模空间是非紧的，在这种空间的积分需要特别小心和额外的定义在玻色弦清，当模空间上的积分发散时，（314）的积分结果只能确定到一定的“边界项”（退化的黎曼曲面或低亏格曲面的贡献）在超弦情形，我们陷入更严峻的问题，因为“模空间边界”这个概念没有定义实际上，在Grassmann奇变量上的积分并不“知道”边界项是什么正因为这一点，费米弦中的积分度量没有好的定义，而且依赖于“规范选择”或对作用量（323）的“零模式”的特殊选择对两圈贡献来说，这个问题可以“经验地”解决（见[88，22]），但在一般情况下，超弦微扰理论在数学上没有很好定义而且，这些还不是形式的问题：同样的障碍也出现Green和Schwarz[91]的不那么几何的方法中。

[130]G.T.Horowitz and J.Polchinski，“Giuge/gravity duality，”gr-qc/0602037.将发表在Toward, Quantum Gravity, ed.Daniele Oriti, Cambridge University Press

[131]http://golem ph.utexas.edu/～distler/blog/archiv/000404.html

[132]IrvingJanis, Victims.ofGroupthink：APsychologicalStudyofForeign-PolicyDeci-sionsandFiascoes（Boston：HoughtonMifoin，1972），p9。当然，这种现象很老了影响卓著的经济学家JohnKennethGalbraith称它为“惯性智慧”他的意思是，“很多意见虽然未必有什么道理，但往往把握在富人和名人的手里，只有莽汉和傻瓜才会拿自己的事业去冒险反对它们。”（引自一篇书评，见FinancialTimes, Aug 12，2004）

[133]IrvingJanis, CruciallDecisions：Leadershipin PolicymakingandCrisisManagement（NewYork：FreePress，1989），p.60.

[134]http://oregonstate.edu/instruct/theory/grpthink.html.

[135]另一个例子是，冯?诺伊曼（JohnvonNeumann）在1932年发表了关于不存在量子理论隐变量的错误证明，曾被广泛引用近30年，直到后来玻姆（DavidBohm）发现了隐变量。

[136]费耶阿本德（1924～1994）是哲学的“叛徒”，“科学最恶劣的敌人”（《自然》杂志对他的评论）。《反对方法》（AgainstMethod）是他最著名的作品，出版于1975年。本书的副标题是“无政府主义知识论纲要”，宣扬“什么都可以”，意思是科学无所谓方法，科学与占星术、巫术等东西并没有本质区别。（此书有周昌忠的译本，上海译文出版社，1992）——译者

[137]见Paul Feyerabend, Killing Time：The Autobiography of Paul Feyerabend（Chica-go：Univ.of Chicago Press，1996）.

[138]例如见Karl Popper, The Logic of ScientificDiscovery（New York：Routledge，2002）.

[139]ThomasS.Kuhn, TheStructureof ScientificRevolutions（Chicago：Univ.ofChicago Press，1962）.

[140]ImreLakatos, ProofsandRefutations（Cambridge, U.K.：CambridgeUniv.Press，1976）.

[141]苏斯金在为人存原理辩护时就称它的批评者为Popperazzi，因为他们需要借助某些证伪方法。但认同批评波普尔将证伪作为科学过程的唯一要素是一回事，而宣扬在科学基础上接受一个没有唯一确定的可以证伪或证实的预言的理论，则是另一回事。在这方面，我很自豪自己是波普尔的信徒。

[142]Alexander Marshack, The Roots of Civilization：The Cognitive Beginnings ofMans First Art, Symbol, and Notation（New York：McGraw-Hill，1972）.

[143]D.H.Wolpert and W.G.Macready, No Free Lunch Theorems for Search, Technical Report, Santa Fe Institute, SFI-TR-95-02-010.

[144]Richard P.Feynman，“What is Science，”The Physics Teacher, Sept.1969.

[145]转引自Simon Singh，“Even Einstein Had His Off Days，”New York Times, Jan.2，2005.

[146]例如，见Mara Beller, Quantum Dialogue：The Making ofa Revolution（Chicago：Univ.of Chicago Press，1999）.

[147]Thomas S.Kuhn, The Structure of Scientific Revolutions（Chicago：Univ.of Chicago Press，1962）.

[148]爱因斯坦1944年12月7日给RA.Thorton的信，未发表，耶路撒冷希伯来大学爱因斯坦档案（EA6-574）引自DonHoward，“AlbertEinsteinasaPhillosopherof

Science，”PhysicsToday, Dec.2005

[149]T.Jacobson and L.Smolin，“Nonperturbative Quantum Geometries，”Nucl.Phys.B，299：295～345（1988）.

[150]例如，见L.Crane，“Clock and Category：Is Quantum C-avity Algebraic?”gr-qc/9504038；J.Math.Phys.，36：6180～6193（1995）.

[151]例如，见F.Markopoulou，“An Insiders Guide to Quantum Causal Histories，”hep-th/9912137；Nucl.Phys.B, Proc.Supp.，88（1）：308～313（2000）.

[152]Seth Lloyd, Programming the Universe：A Quantum Computer Scientist Takes on the Cosmos（New York：Alfred A.Knopf，2006）.

[153]我要再次强调，我说的只是那些经过了良好训练，拿到了博士学位的人，而不管那些不了解科学是什么的人或科学骗子。

[154]L.Smolin，“On the Nature of Quantum Fluctuations and Their Relation to Gravita-tion and the Principle of Inertia，”Glass.Quant.Grav.，3：347～359（1986）.

[155]Julian Barbour, The End of Time：The Next Revolution in Physics（New York：Ox-ford Univ.Press，2001）.

[156]D.Finkelstein，“Past-Future Asymmetry of the Gravitational Field of a Point Parti-cle，”Phys.Rev.，110：965～967（1958）.

[157]Antony Valentini, Pilot-Wave Theory of Physics and Cosmology（Cambridge, U.K.：Cambridge Univ.Press, in press）.

[158]下面国家科学基金会（NSF）1995年给NotreDame大学物理学家JamesCush-ig的一封信的部分内容，信中拒绝了他支持量子理论基础研究的建议：这里考虑的问题，即量子力学的哥本哈根诠释与因果的[玻姆]诠释的争论，已经讨论过多年，而在NSF物理学部看来，问题已经解决了。因果解释与证明贝尔不等式的实验不一致。因此……资助……这个领域的研究项目是不明智的。这封信值得注意的地方是，它犯了一个基本的错误，因为那时专家们已经明白因果解释完全符合检验贝尔不等式的实验。顺便说一句，Cushing在将兴趣转向量子理论基础之前是一个成功的粒子物理学家，但NSF还是砍掉了他的资助。

[159]D.Deutsch, Proc.Roy.Soc.A，400：97～117（1985）.

[160]David Deutsch, The Fabric of Reality：The Science of Parallel Universes and Its Im-plications（London：Penguin，1997）.

[161]P.W.Shor，“Polynomial-Time Algorithms for Prime Factorization and Discrete Log-anthms on a Quantum Computer，”quant-ph/9502807.

[162]A.Valentini，“Extreme Test of Quantum Theory with Black Holes，”astroph/041250s.

[163]Alexander Grothendieck.Ré coltes et Semailles，1986，Roy Lisker英译，见www.grothendieck-cirde.org, chapter 2.

[164]这有一个不幸的例外，那就是教授被他年轻的自我吓坏了，而且已经失去了年轻时候的冒险精神，变成了科学保守主义分子，当然不好让这样的人想起他年轻的时候。

[165]例如，见“A Study on the Status of Women Faculty in Science at MIT，” vol. XI, no.4，March 1999，见http://web.mit.edu/fnl/-women/women.html关于科学中的妇女问题，见美国物理学会网页：http://www.aps.org/educ/-cswp/和哈佛大学教师多样性委员会网页：http://www.aps.org/educ/cswp/.

[166]James Glanz，“Even Without Evidence, String Theory Gains Inouence，”New York Times, March 13，2001.

[167]Gary Taubes, Nobel Dreams：Power, Deceit and the Ultimate Experiment（New York：Random House，1986）pp.254 ~ 255.

[168]对辛格的访谈，发表在http://www.abelprisen-.no/en/prisvinnere/2004/inter-view 20041.html.

[169]对康尼斯的访谈，见www.ipm.ac.ir/IPM/news/connes-interview.pdf

[170]BebelGilberto是来自巴西的流行女歌星，Esthero是多伦多的一个二人组合，RonSexsmith是加拿大有名的创作型歌手